INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University Microfilms International

300 N. ZEEB ROAD, ANN ARBOR, MI 48106
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND

8016068

ZEE, YUN-YANG

PHONETIC STUDIES OF CHINESE TONES

University of California, Los Angeles PH.D. 1980

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106 18 Bedford Row, London WC1R 4EJ, England

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs _____
2. Colored illustrations _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Print shows through as there is text on both sides of page _____
6. Indistinct, broken or small print on several pages ✓ throughout _____
7. Tightly bound copy with print lost in spine _____
8. Computer printout pages with indistinct print _____
9. Page(s) _____ lacking when material received, and not available from school or author _____
10. Page(s) _____ seem to be missing in numbering only as text follows _____
11. Poor carbon copy _____
12. Not original copy, several pages with blurred type _____
13. Appendix pages are poor copy _____
14. Original copy with light type _____
15. Curling and wrinkled pages _____
16. Other _____

University Microfilms International

300 N. ZEEB RD., ANN ARBOR, MI 48106 (313) 761-4700

UNIVERSITY OF CALIFORNIA

Los Angeles

Phonetic Studies of Chinese Tones

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Linguistics

by

Yun-Yang Zee

1980

The dissertation of Yun-Yang Zee is approved.

Hung-Hsiang Chou

Ian Maddieson

Peter N. Ladefoged, Committee Chairman

University of California, Los Angeles

1980

For

Evelyn and Jennifer

Table of Contents

ACKNOWLEDGMENT

I owe my largest debts to Peter Ladefoged. Without him I would not have been able to complete my dissertation. I am grateful to Ian Maddieson, Jean-Marie Hombert and Steve Greenberg for their generous assistance. I appreciate the interaction with George Bedell, Ron Carlson, Leon Jacobson, Gogol Mardirussian, Lloyd Rice, Marcel van den Broecke, Vincent van Heuven and Jim Wright. I treasure the friendships of Cathe Browman, Pat Coady, Sandy Disner, Vivian Flores, Jack Gandour, Manual Godinez, Hector Javkin, Haruko Kawasaki, Mona Lindau, Wendy Linker, Jonas Nartey, George Papcun, Carol Riordan, Mary Smith and Ann Wingate. Special thanks go to Vimal Graham, Willie Martin, Anna Meyers and Renee Wellin for their ever available helping hand.

The chapters in this dissertation have all been reproduced from camera-ready copy prepared for publication. I would like to thank Ian Maddieson for allowing me to include Section 6 of which he is a co-author in the dissertation. The research for the dissertation was supported in part by the National Science Foundation.

VITA

Born January 1st, 1943, Shanghai, China

1966 B.A., Hong Kong Baptist College

1967 M.A., University of Pittsburgh

1972 M.A., Claremont Graduate School

1975 M.A., University of California, Los Angeles

1977-80 Graduate Research Assistant, Phonetics Laboratory, Department of Linguistics, University of California, Los Angeles [funded by National Science Foundation]

PUBLICATION

Zee, Yun-Yang

1977 'Effect of F_0 on the duration of [s]' *UCLA Working Papers in Phonetics* 36, 122-127.

1978 'Duration and intensity as correlates of F_0' *Journal of Phonetics*, 6, 213-220.

1980 'Tone and vowel quality' *Journal of Phonetics*, 8 (to appear)
'Intraoral air pressure of [p] as a function of F_0' *UCLA Working Papers in Phonetics* 49 (to appear)
'Effect of aspiration on fundamental frequency in Cantonese' *UCLA Working Papers in Phonetics* 49 (to appear)
'A spectrographic investigation of Mandarin tone sandhi' *UCLA Working Papers in Phonetics* 49 (to appear)

1981 'Effect of vowel quality on the perception of post-vocalic nasal consonants' *Journal of Phonetics*, 9 (to appear)

Zee, Yun-Yang and Ian Maddieson

1980 'Shanghai tone and tone sandhi: phonetic evidence and phonological analysis' *Glossa* (to appear)

Zee, Yun-Yang and Steven Greenberg

1979 'Perception of contour tones' *UCLA Working Papers in Phonetics* 45, 150-164.

ABSTRACT OF THE DISSERTATION

Phonetic Studies of Chinese Tones

by

Yun-Yang Zee

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 1980

Professor Peter N. Ladefoged, Chairman

This dissertation covers three areas of phonetic study of Chinese tones: (1) the interaction between tone and vowel, (2) the interaction between tone and consonant, and (3) acoustical analysis of tone sandhi. Five major Chinese dialects, Standard Mandarin, Southwestern Mandarin, Shanghai, Taiwanese and Cantonese, were used for investigation. The material is organized in seven sections.

Section 1 investigates whether tone correlates with both duration and intensity in Taiwanese. Results show that (a) the high tone has an overall higher intensity than the mid tone and the mid tone has an overall higher intensity than the low tone, (b) the high-falling tone has an overall higher intensity than the low-rising tone, and (c) the rising tone is longer than the level tone and the level tone is longer than falling tone.

Section 2 investigates the effect of tone on vowel quality in Taiwanese. Results show that (a) vowels [i], [e], [ɔ] and [u], not however [a], associated with the high tone occupy areas in the acoustical vowel space that are distinct from those occupied by the same vowels associated with the low tone, and (b) the average fundamental frequency of the high tone correlates with vowel height. No such correlation is found in vowels produced with the low tone.

Section 3 investigates the difference between the effects of [pʰ] and [p] on the fundamental frequency onset of the following diphthong [ei] in Cantonese. Results show that the F_0 onset for [ei] following [ei] is higher, however, the intensity onset is higher for [ei] following [p], which implies that a higher F_0 may be produced with a decreased subglottal pressure.

Section 4 investigates whether the duration of the prevocalic [s] is influenced by the tone on the following vowel and to examine whether temporal compensation exists between the duration of the fricative and the duration of the vowel within a certain tone class. Results show that (a) the duration of the fricative is influenced by the tone on the following vowel, and it varies according to the onset value of the tone, and (b) within each tone class, temporal compensation does not exist between the durations of the segments [s] and [i] in the syllable [si].

Section 5 investigates whether the peak intraoral air pressure in a voiceless unaspirated stop [p] is affected by the tone on the following vowel or diphthong in Mandarin, Chungking, Shanghai and Cantonese. Results show that (a) there is a good correlation between the peak intraoral air pressure in [p] and the F_0 onset of the tone on the following vowel or diphthong, and (b) the amount of change in subglottal pressure cannot account for the large variation of F_0.

Section 6 investigates whether the pitch contours on monosyllabic words and on compounds are actually similar. The results show that tone spreading will partially explain the patterns of tone contours, on the compounds, but that there are also some more arbitrary processes involved in tone sandhi in Shanghai. A phonological analysis is presented which regards contours as sequences of level tones, and tones are treated as having an 'autosegmental' association with syllabic units.

Section 7 is a spectrographic investigation of the tone sandhi at the phrase and sentence levels in Mandarin. Results show that none

of the three claims proposed by Chao, 1948, 1968 and Cheng, 1968 applies to tone sandhi at the productive or acoustic level in the speech of today's young generation of Peking Mandarin speakers.

Section 1 *Duration and Intensity as Correlates of F_0*

1. Introduction

Ferrein (1741) was probably the first to conduct an experiment on the correlation of pitch with air pressure. He connected the trachea of an excised larynx of a dog to a bellows, and used threads and weights to manipulate the positions of the arytenoid cartilages to create different vocal cord tensions. A manometer was used to measure air pressure. With constant air pressure, pitch increased as he increased the tension of the vocal cords. With increases in air pressure further increases in pitch were produced. That increase in subglottal pressure, or increase in vocal cord tension, increases pitch has been confirmed by more recent studies (Van den Ber, 1957; Ladefoged, 1963; Stevens, 1973, 1975). Ohala and Ewan (1972) reported that, given a comparable fundamental frequency interval, there is a marked tendency for fundamental frequency lowering to be faster than fundamental frequency raising. The present study further investigates these correlations by using a tone language. More precisely, it is concerned with the correlation of fundamental frequency with both duration and intensity. The tone language chosen is Taiwanese which is ideal for investigation as it has high, mid, low, high-falling and low-rising tones. Of the two previous instrumental studies on tones in Taiwanese, Chiang's spectrographic analysis (1967) presents the phonetic shapes of the tones in Taiwanese but it does not deal with the correlation of fundamental frequency with either duration or intensity; Weingartner (1970) analyzes the Taiwanese tones in terms of duration and intensity, but unfortunately he does not provide information regarding his experimental procedures, the instruments he used and his statistical method (for a review of Weingartner, 1970, see Cheng, 1972).

2. Procedure

The five Taiwanese tones used for investigation are high, mid, low, high-falling and low-rising. All these tones occur on the syllable [si]:

high	[sí]	'silk'
mid	[sī]	'yes'
low	[sì]	'four'
high-falling	[sî]	'to die'
low-rising	[sǐ]	'time'

A word list containing 75 tokens of these [si] words (5 words × 15 repetitions) was prepared. Two male native Taiwanese speakers (both university students) were instructed to read the list at a normal rate of speech. The test word was placed in a carrier frame:

[guá kóŋ ___ tài kēi tʰī̂a]

'I say _____ everyone listens'

The recording was made in a sound treated room at a single session for each speaker. The tapes were analyzed, using a PDP-12 computer at the UCLA Phonetics Laboratory. Fundamental frequency measurement was obtained every 10 msec, using the Cepstrum method. RMS amplitude was also taken every 10 msec, using a square window. Duration measurements were made on the waveform displayed on the computer screen.

Since utterances have different lengths, it is necessary to select comparable points at which the differing fundamental frequencies may be compared. Accordingly, the values of fundamental frequency and RMS amplitude for each token were divived into five parts, each containing as nearly as possible equal numbers of 10 msec interval. The fundamental frequency and RMS values of each part were then taken to be the means of the sets of 10 msec intervals within that part. The 15 repetitions of each syllable were then averaged so as to give mean values for each of the five parts. Thus, the first point on any curve in Figure 1 and Figure 2 represents the mean of the mean values for the initial time intervals of 15 repetitions of that syllable. The second point is the mean of the mean for the second set of time intervals and so on. Each bar in Figure 3 and Figure 4 represents the mean of 15 tokens for each tone class. The intensity value of the low tone was taken as reference (0 db) to compute the intensity values of other tones.

3. Results

3.1. Intensity

As shown in Figure 1 and Figure 2, there is a good correlation between fundamental frequency and intensity levels. In general, the intensity curves fall and rise with the fundamental frequency curves. However, we can see that the onset points in the intensity curve for both the high and the mid tones do not correlate so well, although, as shown in Figure 3 and Figure 4, the average RMS value is still correlated with the average fundamental frequency for both speakers as the high tone has a higher overall intensity than the mid tone and the mid tone has a higher overall intensity than the low tone. As far as the contour tones are concerned, the intensity curves again correlate closely with the direction of fundamental frequency change, although the offset points on the intensity curve for the low-rising tone do not correlate so well for both speakers. For these contour tones, as shown in Figure 3 and Figure 4, the average RMS value is also correlated with the average fundamental frequency as the high-falling tone has a higher overall intensity than the low-rising tone.

3.2. Duration

As shown in Figure 3, for Speaker 1, the high tone has a longer duration than the mid tone, and the mid tone has a longer duration than the

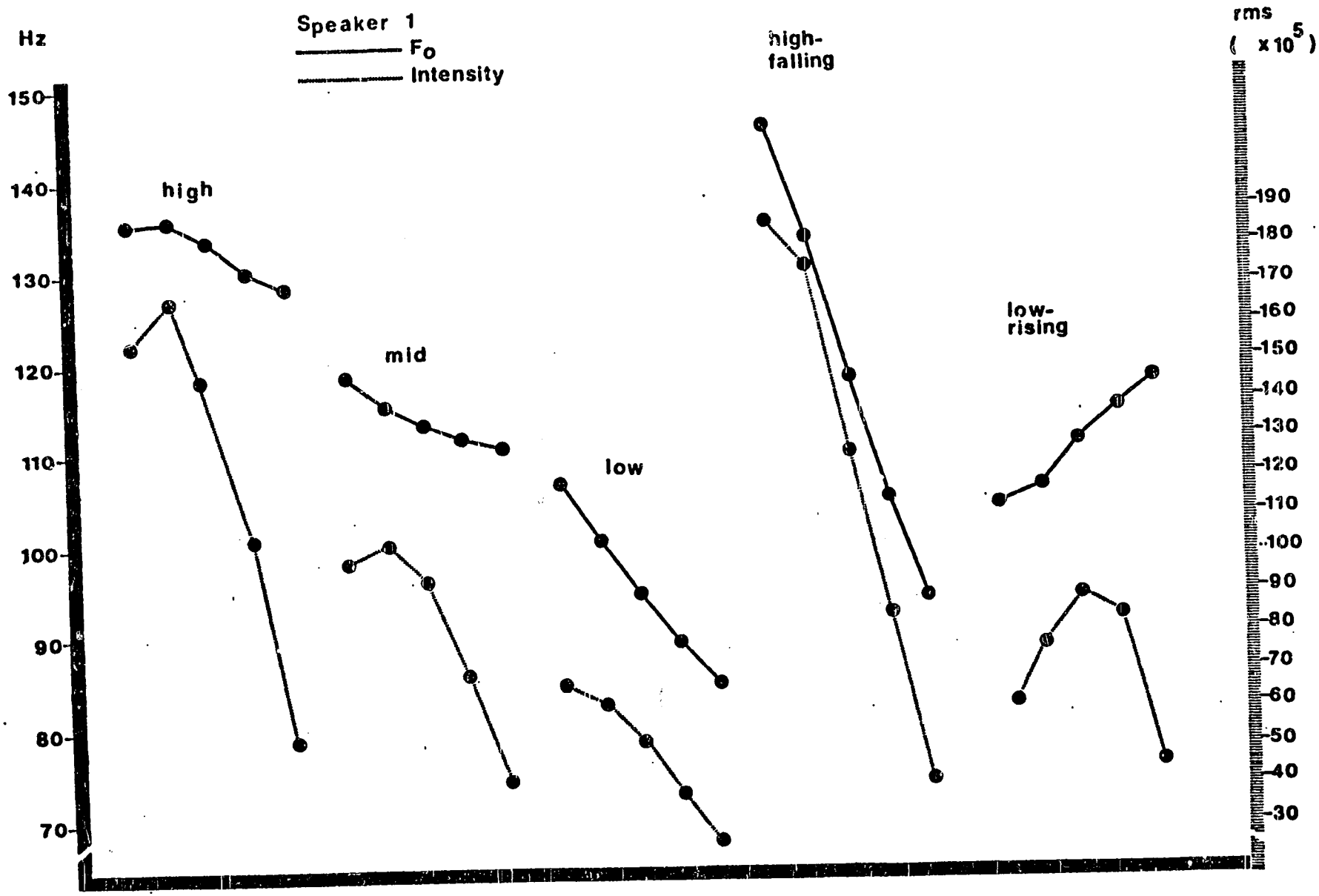


Figure 1

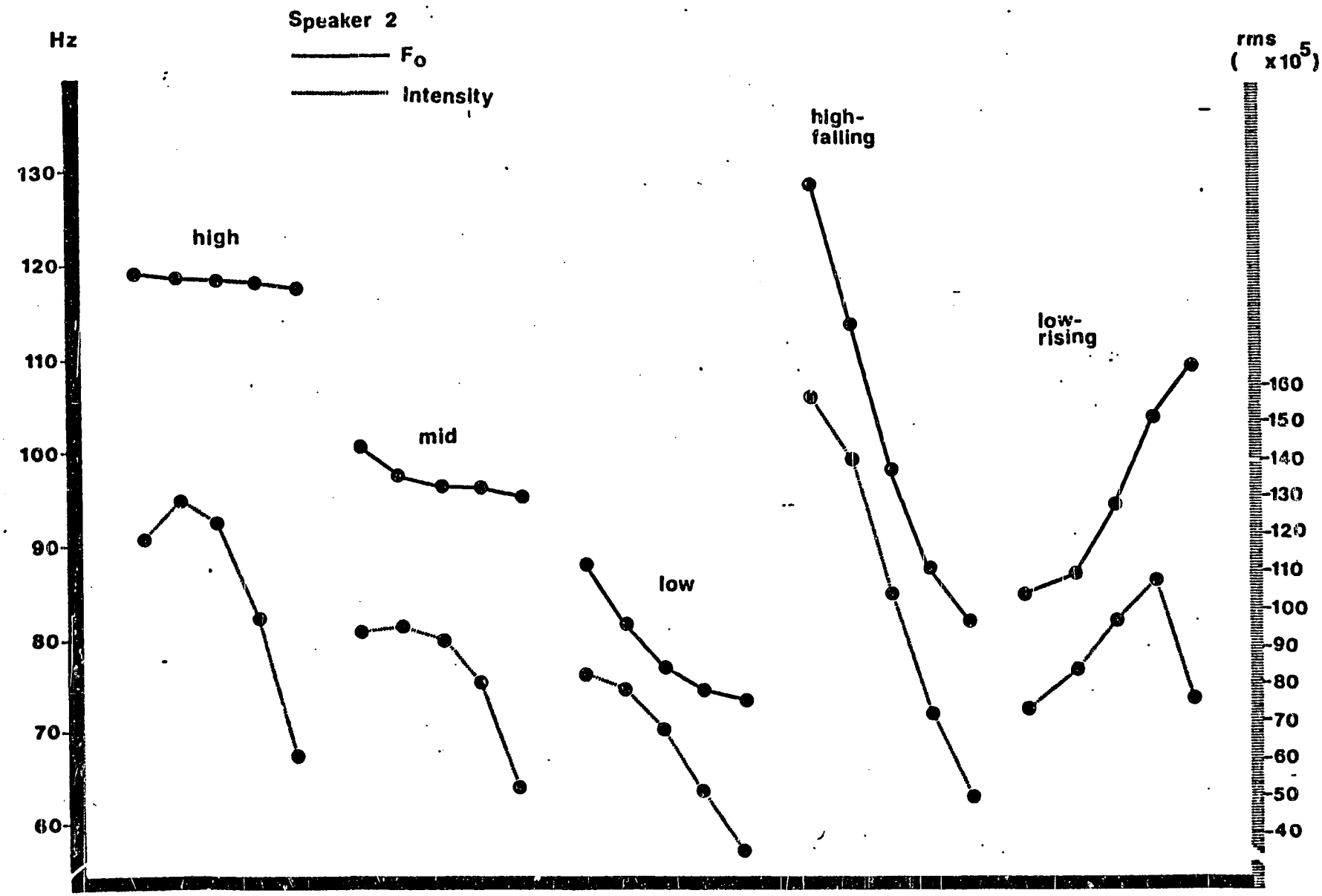


Figure 2

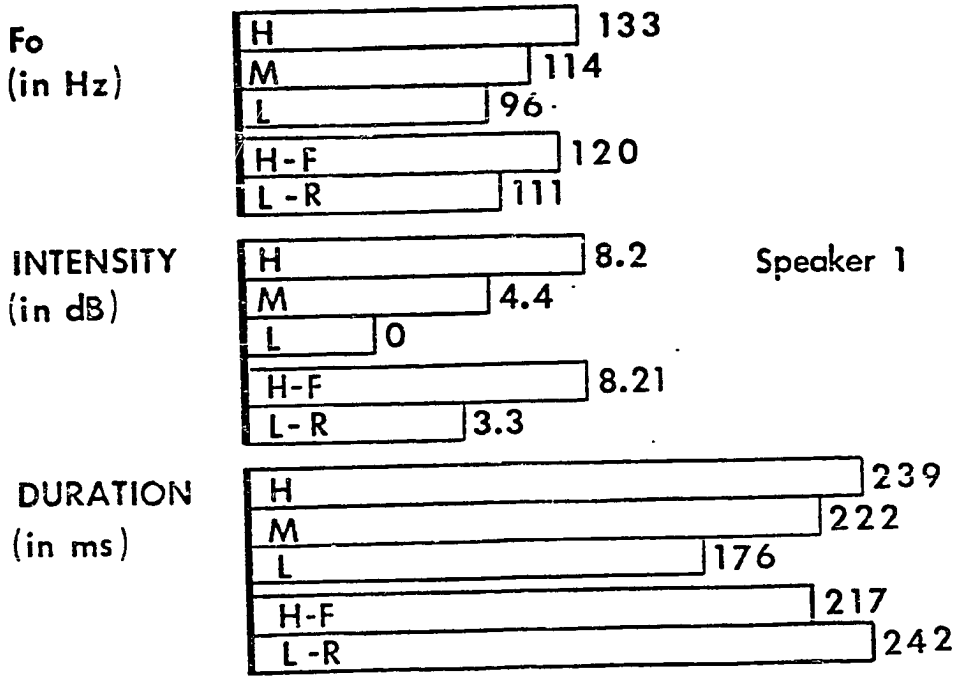


Figure 3

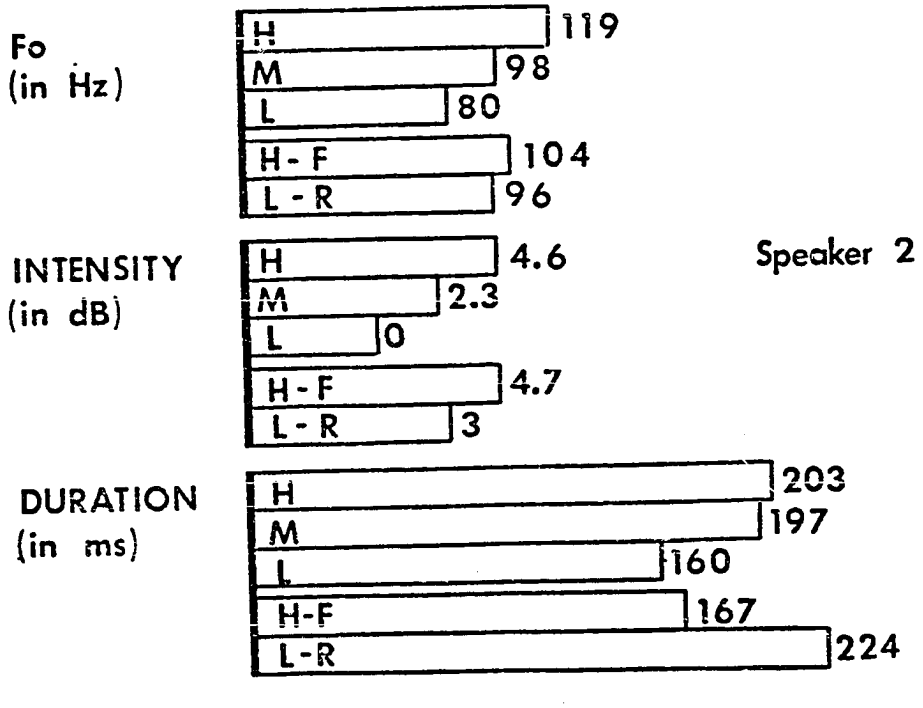


Figure 4

low tone. The low-rising tone has a longer duration than the high-falling tone. T-test results reveal that:
(a) the difference in duration between the high and the mid tone is non-significant;
(b) the difference in duration between the high and the low tones and between the mid and the low tones are significant at the .01 level for both speakers;
(c) the difference in duration between the low-rising tone and the high-falling tone is significant at the .01 level.
Table I shows the results of grouped data T-test for durations of the different tones for Speaker 1.

Similar results were also found in Speaker 2 as shown in Figure 4. The high tone has a higher overall intensity and longer duration than the mid tone, and the mid tone has a higher overall intensity and longer duration than the low tone. The high-falling tone has a higher overall intensity than the low-rising tone, although the low-rising tone has a longer duration. Table II shows the results of grouped data T-test for durations for Speaker 2.

3.3. The low tone as a low-falling tone

It should be noticed that the low tone may be considered a low-falling tone, as the fundamental frequency range covered by the low tone is substantially greater than that covered by the high or the mid tone for both speakers as shown in Table III.

For Speaker 1, the fundamental frequency range difference is 7 Hz for the high tone, 7 Hz for the mid tone and 21 Hz for the low tone. For Speaker 2, the fundamental frequency range difference is 1 Hz for the high tone, 5 Hz for the mid tone and 14 Hz for the low tone. It seems appropriate to consider the high and the mid tones the level tones and the low tone the low-falling tone.

3.4. The high-falling tone vs. the low-rising tone

Figure 1 and Figure 2 show that the high-falling tone covers a much greater range than the low-rising tone: 52 Hz for the high-falling tone vs. 14 Hz for the low-rising tone for Speaker 1, and 46 Hz for the high-falling tone vs. 24 Hz for the low-rising tone for Speaker 2. Since the duration of the high-falling tone is shorter than that of the low-rising tone for both speakers (Figure 3 and Figure 4), the rate of change of fundamental frequency is thus substantially greater for the high-falling tone than for the low-rising tone for both speakers. For Speaker 1, the linear rate of fundamental frequency change is 235 Hz/sec in the high-falling tone and 57 Hz/sec in the low-rising tone. For Speaker 2, the linear rate of change in the high-falling tone is 275 Hz/sec and 152 Hz/sec in the low-rising tone. These are shown in Table IV.

4. Discussion

4.1. That the high tone has a higher overall intensity than the mid tone and the mid tone has a higher overall intensity than the low tone may be because higher frequency correlates with both greater tension of vocal cords and higher subglottal pressure. The vocal cords will be driven by a heightened subglottal pressure which will lead to higher intensity. Increased subglottal pressure (Ladefoged, 1963) and increased tension of the vocal cords (Stevens, 1973, 1975) have been considered as separate factors, either one of which acting alone is capable of producing an increase of fundamental frequency. In natural speech increased subglottal pressure and increased tension of the vocal cords may operate together to get maximum efficiency. An increase of each may be capable of producing a great increase in fundamental frequency. However, there are cases when increased tension of the vocal cords must operate alone to increase fundamental frequency. As we have indicated in 3.1., the onset points of the intensity curve for both the high and the mid tones and the offset points of the low-rising tone do not correlate so well with the directions of fundamental frequency change. In fact, in both the high and the mid tones, the second half of the intensity curve falls sharply although the fundamental frequency curve falls in a much lesser degree for both speakers (Figure 1 and Figure 2). These indicate that a higher fundamental frequency may be produced without a corresponding higher subglottal pressure. Particularly in the case of the low-rising tone where the intensity curve falls sharply while the fundamental frequency curve rises. Thus, increased tension of the vocal cords alone, perhaps with the assistance of larynx elevation, is capable of producing a higher fundamental frequency.

4.2. For both speakers, the low-rising tone is longer than the high-falling tone and the range covered by the high-falling tone is greater than the range covered by the low-rising tone. As can be seen in Tables III and IV, the rate of change of fundamental frequency is greater for the high-falling tone than for the low-rising tone. The reason may be that it requires more time for the vocal cords to increase the rate of vibration than to decrease it. This observation is supported by the result obtained by Ohala and Ewan (1972) that there is a marked tendency for pitch lowering to be faster than for pitch raising for a comparable pitch interval. Similarly, Sundberg (1973) has demonstrated that unlike singers, untrained subjects perform pitch drops considerably faster than pitch elevations. Ohala's explanation for this phenomenon is that it requires more physiological effort to raise a pitch than to lower it. As shown in Figure 1 and Figure 2, the intensity curve falls sharply at the offset point, whereas the fundamental frequency curve rises, that is, fundamental frequency rises without the presence of a higher subglottal pressure. Perhaps it is the lack of an accompanying higher subglottal pressure that slows down the speed of fundamental frequency change in the low-rising tone. It is likely that the extra physiological effort that Ohala refers to is needed when the subglottal pressure begins to drop during the process of fundamental frequency raising. To compensate for the loss of acoustic energy, such extra physiological effort is

Table I. Results of grouped data T-test for durations (in msec) of (a) high and mid tones, (b) high and low tones, (c) mid and low tones, and (d) high-falling (H-F) and low-rising (L-R) tones (Speaker 1)

	(a)		(b)		(c)		(d)	
	High	Mid	High	Low	Mid	Low	H-F	L-R
MEAN	243	229	243	176	229	176	218	246
STD DEV	28	17	28	14	17	14	19	27
T-SCORE	1.63		8.22		9.34		3.12	

Table II. Results of grouped data T-test for durations (in msec) of (a) high and mid tones, (b) high and low tones, (c) mid and low tones, and (d) high-falling (H-F) and low-rising (L-R) tones (Speaker 2)

	(a)		(b)		(c)		(d)	
	High	Mid	High	Low	Mid	Low	H-F	L-R
MEAN	203	196	203	159	196	159	166	224
STD DEV	18	14	18	20	14	20	17	14
T-SCORE	1.07		6.07		5.63		9.73	

Table III. Fundamental frequency (in Hz) ranges of high, mid and low tones and their range differences (Speaker 1 and Speaker 2)

	Speaker 1		Speaker 2	
	F_o range	Difference	F_o range	Difference
High	135-128	7	119-118	1
Mid	118-111	7	101- 96	5
Low	107- 86	21	88- 74	14

Table IV. Fundamental frequency (in Hz) ranges and linear rate of fundamental frequency change in high-falling and low-rising tones (Speaker 1 and Speaker 2)

	Speaker 1		Speaker 2	
	F_o range	Linear rate of F_o change	F_o range	Linear rate of F_o change
High-falling	54	235 Hz/sec	46	275 Hz/sec
Low-rising	14	57 Hz/sec	24	152 Hz/sec

Table V.

	Conditioning		Phonetic realization
intrinsic duration/ intrinsic intensity of a vowel	tone 1	→	duration/intensity 1
	tone 2	→	duration/intensity 2
	tone 3	→	duration/intensity 3
	tone 4	→	duration/intensity 4
	tone 5	→	duration/intensity 5
	.		.
	.		.
	.		.

needed for further increasing the tension of the vocal cords and further raising the larynx in order that the fundamental frequency may be raised.

4.3. That the high tone and the mid tone have a longer duration than the low tone may be because the low tone is actually a low-falling tone as indicated in 3.3. Statistical results have shown that the difference in duration between the high and the mid tones is non-significant, which may be due to the fact that they are both level tones. It seems that the difference in duration between different tone classes is primarily determined by the shape, rather than the level, of the fundamental frequency curve. As the level (high and mid) tones are slightly falling, we would expect the duration of these level tones to come in between those of the low-rising tone and of the falling (high-falling and low-falling) tones. As a matter of fact, for both speakers, the level tones are shorter than the low-rising tone and longer than the falling tones.

4.4. Lehiste refers to intrinsic duration as duration of a segment as determined by its phonetic quality (1970:18); and to intensity of a vowel as intensity considered in relation to its phonetic quality (1970: 120). These definitions are only partially applicable to a tone language, such as Taiwanese, for in a tone language the phonetic realization of a vowel must include both its segmental quality and its relative pitch or tone. It may be than that the 'intrinsic duration and intensity' of a tone interacts with the intrinsic duration and intensity constrained by the vowel quality, or that the intrinsic duration and intensity of a vowel in a tone language is determined by its phonetic quality as further conditioned by its tone. (See Table V.)

It seems then the intrinsic duration or the intrinsic intensity of a vowel in a tone language always remain recessive and will only manifest itself as a variant.

5. Summary of results

The purpose of this experiment was to determine whether fundamental frequency correlates with both intensity and duration. A tone language, Taiwanese, was used for investigation. The results from analyzing the data produced by two male Taiwanese speakers and discussion of these results indicate:

(a) The high tone has an overall higher intensity than the mid tone and the mid tone has an overall higher intensity than the low tone.

(b) The high tone and the mid tone have a longer duration than the low tone. The high tone has a longer duration than the mid tone, but the difference between them is non-significant.

(c) The high-falling tone has an overall higher intensity than the low-rising tone, although the low-rising tone has a longer duration.

(d) If the low tone is considered a low-falling tone, the low-rising tone is longer than the level (high and mid) tones and the level tones are longer than the falling (high-falling and low-falling) tones.

(e) The linear rate of fundamental frequency change is greater for the

high-falling tone than for the low-rising tone.
(f) The high-falling tone covers a greater range of fundamental frequency than the low-rising tone.
(g) Regardless of the tone classes, the intensity tends to drop at the offset point of a tone.
(h) The difference in duration between different tone classes is primarily determined by the shape of the fundamental frequency contour.
(i) The intrinsic duration and the intrinsic intensity of a vowel in a tone language are conditioned by the tone that the vowel carries.

Acknowledgment

I would like to thank the members of UCLA Tone Seminar for many helpful comments. Special thanks are owed to Jean-Marie Hombert, Peter Ladefoged, Victoria Fromkin and Steve Greenberg. This research was supported by NSF.

References

Chen, T.M. (1972) Review of Tone in Taiwanese, an instrumental investigation, by F.F. Weingartner. *Lingua* 30: 2, 3, 285-294.

Chiang, H.T. (1967) Amoy-Chinese tones. *Phonetica* 17, 100-115.

Ladefoged, P. (1963) Some physiological parameters in speech. *Language and Speech* 6, 109-119.

Lehiste, I. (1970) *Suprasegmentals* Cambridge, Massachusetts: M.I.T. Press.

Ohala, J. and W. Ewan (1972) Speech of pitch change. Paper presented at the 94th ASA-Meeting, Autumn, 1972, Miami Beach, Florida.

Stevens, K.N. (1973) Theory of vocal cord vibration and its relation to laryngeal features. Paper presented at the 86th ASA-Meeting, Autumn, 1973, Los Angeles, California.

Stevens, K.N. (1975) Physics of laryngeal behavior and larynx modes. Paper presented at the Seminar on Larynx and Language, International Congress of Phonetics, August, 1975, Leeds, England.

Sundberg, J. (1973) Data on maximum speed of pitch changes. *STL-QPSR*, 4/1973, 39-47.

Van den Berg, J. (1957) Subglottalic pressures and vibrations of the folds. *Folia Phoniatrica* 9, 65-71.

Weingartner, F. (1970) Tones in Taiwanese, an instrumental investigation. Taipei, Taiwan: National Taiwan University, College of Arts, Monograph No. 2.

Section 2 *Tone and Vowel Quality*

1. INTRODUCTION

It has been widely attested that the production of tones often involves the movement of the larynx, and also that a higher pitch is correlated with a higher larynx position. Parmenter et. al. (1933) conducted an X-ray investigation of the changes which take place in the articulation of a vowel when it is pronounced first on a low pitch and then on a high pitch. Three speakers, speaking English, French and Spanish respectively, were used. They concluded that important modifications occur in the articulation of a vowel when its pitch is raised an octave. These modifications include (1) the raised larynx, (2) changes in the supraglottal cavities which are probably due to the raised larynx, and (3) changes in these cavities which are probably not due to the raised larynx. The correlation of higher pitch with higher larynx height has been further investigated recently by phoneticians, using different techniques, such as Vanderslice's cricothyrometer (1967) and Ewan and Krones' thyroumbrometer (1972, 1974). Results similar to that of Parmenter's were reported. It is quite obvious that a raised or lowered larynx will alter the length of the pharynx and thus the length of the entire vocal tract, leading to a change in formant frequencies. Based on the assumption that a raised or a lowered larynx will shorten or lengthen the pharynx, Sundberg and Nordström (1976) investigated the effect of raised larynx on vowel formant frequencies. Two subjects, a phonetician and a singer, were used in their experiment. Both were able to control fairly well the positioning of their larynges. During phonation, the subjects were instructed to keep a finger on the thyroid cartilage to check the larynx position. Their findings in regard to the effects on the formant frequencies of a raised larynx were: (1) a substantial rise in the second formant frequency in high front vowels, (2) a rise in both the first and the second formant frequencies in open vowels, and (3) a rise in several vowels in both the third and the fourth formant frequencies.

The present study investigates the effect of tone on vowel quality by using a tone language. It differs from Sundberg and Nordström's in that it does not attempt to isolate pharyngeal change as a single factor that affects formant frequencies. Thus, speakers were not instructed to consciously lower or raise their larynges during phonation or to consciously sustain a vowel, nor were they instructed to keep a finger on the thyroid cartilage. The study explores how formant frequencies change when a vowel is pronounced with different tones in natural speech produced by speakers of a tone language.

2. PROCEDURE

Five Taiwanese Chinese vowles [i], [e], [a], [ɔ] and [u] were used for investigation. These vowels were uttered with two contrasting tones,

which are the high tone and the low tone. In most cases, these vowels, associated with either the high or the low tone, are meaningful monosyllabic morphemes in this dialect, as shown below:

	High tone	Low tone
[i]	'he'	'meaning'
[e]	'to nudge'	'to strangle'
[a]	'Asia'	'secondary'
[ɔ]	'dark'	'bay'
[u]	'mud'	'dirty'

A word list containing these vowels, 10 repetitions for each tone class, was prepared. These test words were arranged in a random order and each vowel was placed in the carrier frame as shown below:

[guá	tɕí-má	tʰà?	___	tài-kēi tʰía]
I	now	read	___	everyone listen
'I	now	read	___	everyone listens'

Three male native Taiwanese speakers, all in their late twenties and all from southern Taiwan were instructed to produce these tokens at a normal rate of speech. The recording was done in a single session for each speaker in a sound treated room. The data was analyzed by using a PDP-12 computer at the UCLA Phonetics Lab. Formant frequency values were obtained from LPC spectra and fundamental frequency from the corresponding Cepstrum. For each vowel, values were obtained for five tokens with a high tone, and five tokens with a low tone. In some cases, additional tokens were analyzed when the LPC did not give complete formant frequency values for a vowel. The LPC has a window size of 25.6 msec and formant frequency values were obtained every 10 msec. Each vowel associated with either the high or the low tone was analyzed entirely, that is, from the beginning to the end of its waveforms. The formant frequency values for each 10 msec interval were averaged for each token. The averaged value which was taken to be the formant frequency (F1, F2, or F3) value for a vowel is shown in Tables, I, II and III for Speakers 1, 2 and 3. The reason for taking the average formant frequencies for each token is because there is an interdependency of formant frequencies and harmonics (Fant, 1973). The dominating harmonics have an influence on the formant frequencies, as indicated in Figure 1 which shows the F_0 and F1 contours as well as the 5th harmonics for vowel [e] produced with a low tone by Speaker 1. Notice how F1 fluctuates particularly when F_0 is high due to the influence of the harmonics. Thus, the value at any one time point for the apparent F1 does not necessarily represent the true value of the formant frequency. Fant (1973) suggests a method of visual interpolation, which is plausible in many circumstances. However, in a long vowel in which F_0 is changing, especially the low tone, it is more accurate to average the values across all the time points. This will of course average out any real variations in formant frequency that are due to changes

in the vocal tract configuration during the production of the vowel. But, in the case of the present study, this will merely support the null hypothesis (that are no variations in formant frequency associated with variations in tone) and thus will not lead us to any false conclusions.

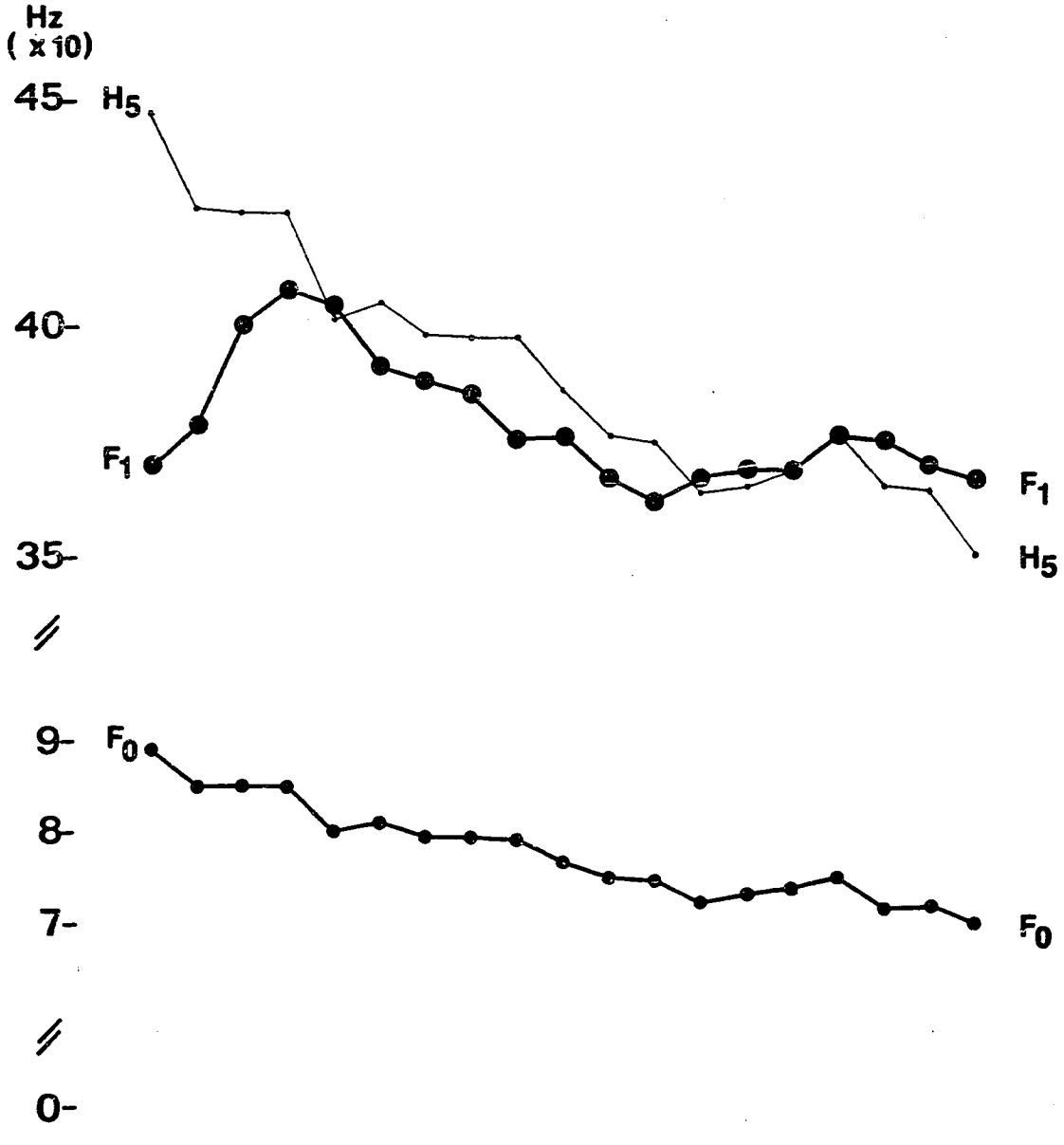


Figure 1. Vowel [e] spoken on a low tone by Speaker 1, showing variations in formant frequency due changes in F_0

It should also be noticed that it is often impossible for the LPC analysis to extract certian formant frequencies for certain vowels. For all three speakers, the second formant frequency for the vowels [a], [u] and [ɔ] often could not be tracked by the LPC. Ladefoged (1967), analyzing formant frequencies using a sound spectrograph, encountered similar difficulties. He found that 'both [ɔ] and [u] often (39 per cent of the time) could not be analysed in terms of three formants, because it was impossible to locate the position of formant two; and the complete analysis of [u] was usually (84 per cent of the time) impossible. In the case of these vowels as with [a], it would seem unlikely that any formant tracking machine would be any more successful in locating the formant frequencies' (p. 90). To solve the problem, the second difference of the Linear Prediction Spectrum (LPC2D) was used. The LPC2D was able to locate the second formant frequencies for most of the tokens, except for the vowel [u] produced with the low tone by Speaker 2, where it was successful on four (out of 10) tokens only. The LPC2D has a window size of 51.2 msec and the values of formant frequencies were obtained every 25.6 msec. Due to the different formant frequency analysis methods, the vowels [i], [e] (by LPC) are expected to occupy slightly different areas in the acoustic vowel space in relation to the vowels [a], [u] and [ɔ] (by LPC2D). This difference however is not particularly relevant for our study, since we are concerned with the locations of each individual vowel produced with different tones.

3. RESULTS

The results of analysis for Speakers 1, 2 and 3 are shown in Tables I, II and III respectively. They show the values of F1, F2 and F3 for each vowel associated with the high or the low tone. Also shown are the average fundamental frequency for the high or the low tone for each vowel. There are five values for each category (F_0, F1, F2 or F3), representing five different tokens.

Tables Ia, IIa and IIIa show the means and the standard deviations of the values of F1, F2 and F3 for each vowel associated with the high or the low tone for Speakers 1, 2 and 3 respectivley. They also show the means of the values of fundamental frequency and the results of grouped data t-tests for formant frequencies associated with the tones for each vowel.

The level of significance of the results of grouped data t-tests for formant frequencies associated with the high or the low tone for all three speakers are summarized in Table IV.

Figures 2, 3 and 4 show vowel ellipses associated with the high tone and the low tone. For each vowel cluster of 5 tokens, associated either with the high tone or with the low tone, an ellipse was drawn along axes which are oriented along the principal components, with the radii being two stan-

Table I. Values of F_0, F1, F2 and F3 for vowels associated with the high and the low tones for Speaker 1

	[i]		[e]		[a]		[ə]		[u]	
	High	Low	High	Low	High	Low	High	Low	High	Low
F_0	132	885	120	77	119	78	129	83	136	83
	134	78	117	75	123	78	136	81	135	88
	139	78	125	82	120	79	132	84	135	83
	138	81	137	76	124	79	133	89	134	85
	126	83	123	77	118	79	137	85	145	81
F1	277	278	383	379	881	889	533	546	293	290
	273	264	396	375	892	906	518	530	284	282
	275	269	399	378	868	905	532	526	283	291
	274	267	397	377	888	974	508	507	283	293
	263	275	389	376	842	1030	531	523	293	310
F2	2571	2385	2316	2231	1159	1182	787	682	829	602
	2728	2476	2374	2180	1160	1185	835	657	757	581
	2607	2467	2368	2274	1217	1209	731	648	821	582
	2608	2360	2345	2241	1226	1258	760	641	797	628
	2562	2460	2350	2271	1193	1246	820	677	825	609
F3	3294	3438	2744	2713	2351	2419	2591	2635	2606	2521
	3337	3469	2794	2740	2335	2446	2466	2634	2611	2486
	3265	3520	2827	2889	2309	2434	2280	2509	2620	2531
	3383	3350	2727	2869	2284	2421	2351	2463	2587	2441
	3347	3448	2906	2886	2374	2405	2496	2585	2654	2494

Table Ia. Results of grouped data T-test for formant frequency values of vowels associated with the high and the low tones for Speaker 1

Vowel	Tone	F_0 Mean	F_1 Mean	F_1 SD	F_1 T-score	F_2 Mean	F_2 SD	F_2 T-score	F_3 Mean	F_3 SD	F_3 T-score
[i]	High	134	274	5.5	0.51	2616	66.0	4.90	3325	46.2	3.47
	Low	81	271	5.8		2430	55.2		3445	61.8	
[e]	High	124	393	6.6	5.16	2351	22.8	5.60	2799	71.5	0.39
	Low	77	377	1.6		2239	38.0		2819	1.4	
[a]	High	121	874	20.2	2.38	1191	31.2	1.20	2331	35.3	5.48
	Low	79	941	59.4		1216	34.7		2425	15.6	
[ə]	High	133	524	11.0	0.25	787	42.6	6.07	2437	122.9	1.98
	Low	84	526	14.0		661	17.9		2565	76.8	
[u]	High	137	289	5.2	0.78	806	30.0	12.80	2616	24.6	6.29
	Low	84	293	10.3		600	19.7		2495	35.2	

Table II. Values of F_0, F1, F2 and F3 for vowels associated with the high and the low tones for Speaker 2

	[i]		[e]		[a]		[ɔ]		[u]	
	High	Low	High	Low	High	Low	High	Low	High	Low
F_0	159	108	145	100	145	106	152	103	153	105
	154	103	140	102	138	106	148	101	162	106
	158	108	152	111	143	106	143	103	155	104
	153	106	144	108	138	104	152	113	159	105
	151	103	149	96	144	108	149	102	157	107
F1	303	275	406	379	963	946	450	393	310	313
	300	265	410	366	916	960	422	396	315	328
	302	289	408	387	969	976	423	417	318	318
	299	279	393	379	1007	974	431	387	321	324
	297	283	424	374	895	976	430	371	307	326
F2	2244	2307	2047	2197	1273	1242	794	770	753	537
	2176	2342	2036	2178	1233	1226	697	728	778	620
	2218	2309	2077	2307	1308	1267	739	781	777	573
	2223	2289	2139	2203	1254	1225	740	783	806	675
	2209	2249	2162	2329	1168	1267	774	786	748	-
F3	2890	2991	2732	2763	2405	2621	2846	2853	2491	2503
	2942	2984	2709	2680	2468	2619	2790	2814	2515	2598
	3020	3042	2714	2750	2559	2698	2721	2764	2571	2512
	2902	3015	2799	2744	2445	2618	2846	2840	2607	2626
	3031	2950	2732	2798	2554	2640	2780	2815	2451	2456

Table IIa. Results of grouped data T-test for formant frequency values of vowels associated with the high and the low tones for Speaker 2

Vowel	Tone	F_0	F_1			F_2			F_3		
		Mean	Mean	SD	T-score	Mean	SD	T-score	Mean	SD	T-score
[i]	High	155	300	2.4	5.28	2214	24.8	4.53	2957	65.5	1.18
	Low	106	278	9.0		2299	34.0		2996	34.5	
[e]	High	146	408	11.1	5.18	2092	55.9	3.76	2737	36.0	0.39
	Low	103	377	7.7		2243	69.7		2747	42.9	
[a]	High	142	950	44.6	0.82	1247	52.1	0.07	2486	68.0	4.50
	Low	106	967	12.1		1245	20.9		2639	34.0	
[ɔ]	High	149	431	11.3	4.28	749	37.2	1.05	2800	56.9	0.57
	Low	104	392	16.6		770	24.0		2817	34.1	
[u]	High	157	314	5.7	2.02	772	23.2	5.95	2527	62.4	0.28
	Low	105	322	6.2		601	59.8		2539	70.7	

Table III. Values of F_0, F1, F2, and F3 for vowels associated with the high and the low tones for Speaker 3

	[i]		[e]		[a]		[ɔ]		[u]	
	High	Low	High	Low	High	Low	High	Low	High	Low
F_0	176	114	167	120	162	122	167	123	178	123
	184	112	165	121	169	115	175	110	175	122
	185	107	168	111	164	112	167	114	176	121
	178	108	166	116	166	116	170	117	178	114
	180	113	168	117	169	120	170	114	175	120
F1	294	246	400	395	570	655	505	357	349	304
	309	255	389	382	502	675	516	426	343	317
	328	252	397	402	463	684	499	408	344	327
	316	276	411	377	466	668	506	436	349	318
	329	252	395	396	489	655	496	379	338	301
F2	2416	2307	2325	2221	1311	1216	858	552	628	547
	2428	2137	2441	2205	1231	1269	921	619	672	565
	2480	2247	2427	2273	1302	1252	952	577	551	583
	2444	2142	2252	2216	1322	1263	942	625	628	569
	2496	2144	2459	2226	1246	1247	976	568	540	564
F3	3520	3491	3125	3053	2847	2539	2756	2892	2676	2959
	3406	3397	3111	3311	2702	2416	2736	2934	2734	2811
	3463	3416	3142	3239	2830	2508	2873	2793	2639	2879
	3429	3520	3210	3388	2755	2523	2922	2696	2676	2776
	3416	3401	3285	3167	2784	2453	2918	2711	2735	2765

Table IIIa. Results of grouped data T-test for formant frequency values of vowels associated with the high and the low tones for Speaker 3

Vowel	Tone	F_0	F_1			F_2			F_3		
		Mean	Mean	SD	T-score	Mean	SD	T-score	Mean	SD	T-score
[i]	High	181	315	14.4	7.11	2453	34.1	6.80	3448	46.8	0.08
	Low	110	256	11.5		2195	77.5		3445	56.6	
[e]	High	167	410	29.0	1.46	2395	55.4	6.10	3174	72.4	0.86
	Low	117	390	10.5		2228	26.2		3231	129.3	
[a]	High	166	498	43.4	8.38	1282	41.0	1.61	2783	58.4	8.49
	Low	117	667	12.7		1249	20.6		2488	51.6	
[ɔ]	High	170	505	9.3	6.81	930	44.7	13.86	2841	89.1	0.58
	Low	115	401	32.9		588	32.2		2805	106.2	
[u]	High	176	345	4.6	5.97	604	56.3	1.48	2692	41.6	3.59
	Low	116	313	10.7		566	12.9		2838	80.9	

Table IV. Level of significance (two-tailed test) for Speakers 1, 2 and 3

Speaker	Vowel	F1 (High/Low)	F2 (High/Low)	F3 (High/Low)
	[i]	n.s.	0.01 (H>L)	0.01 (H<L)
	[e]	0.001 (H>L)	0.001 (H>L)	n.s.
1	[a]	0.05 (H<L)	n.s.	0.001 (H<L)
	[ɔ]	n.s.	0.001 (H>L)	0.10 (H<L)
	[u]	n.s.	0.001 (H>L)	0.001 (H>L)
	[i]	0.001 (H>L)	0.01 (H<L)	n.s.
	[e]	0.001 (H>L)	0.01 (H<L)	n.s.
2	[a]	n.s.	n.s.	0.01 (H<L)
	[ɔ]	0.01 (H>L)	n.s.	n.s.
	[u]	0.10 (H<L)	0.001 (H>L)	n.s.
	[i]	0.001 (H>L)	0.001 (H>L)	n.s.
	[e]	n.s.	0.001 (H>L)	n.s.
3	[a]	0.001 (H>L)	n.s.	0.001 (H>L)
	[ɔ]	0.001 (H>L)	0.001 (H>L)	n.s.
	[u]	0.001 (H>L)	n.s.	0.01 (H<L)

(H: High tone '>': greater
L: Low tone '<': smaller n.s.: nonsignificant)

dard deviations (Davis, 1977). We can see in Figures 2, 3 and 4 for all three speakers the vowel ellipses associated with different tones do not overlap, except for vowel [a] for Speakers 2 and 3; that is, vowels [i], [e], [ɔ] and [u] associated with the high or with the low tone occupy distinct areas in the acoustic vowel space for all three speakers.

Some observations in terms of the relationship between vowel formant frequencies and tones across speakers can be made as follows:

1. For Speakers 1 and 2, there are significant differences in the same direction in F1 at the 0.001 level for vowel [e]; for Speakers 1 and 3, there are significant differences, but in the opposite directions, in F1 at the 0.05 or 0.001 level for vowel [a].

2. For Speakers 2 and 3, there are significant differences in the same direction in F1 at the 0.10 or 0.001 level for vowels [i] and [ɔ]; for the same speakers, there are significant differences, but in the opposite directions, in F1 at the 0.01 or 0.001 level for vowel [u].

3. Considering all three speakers, for the vowels [i] and [e], there are significant differences in F2 at the 0.01 or 0.001 level, but in each case for one speaker the differences are in the opposite directions to the other

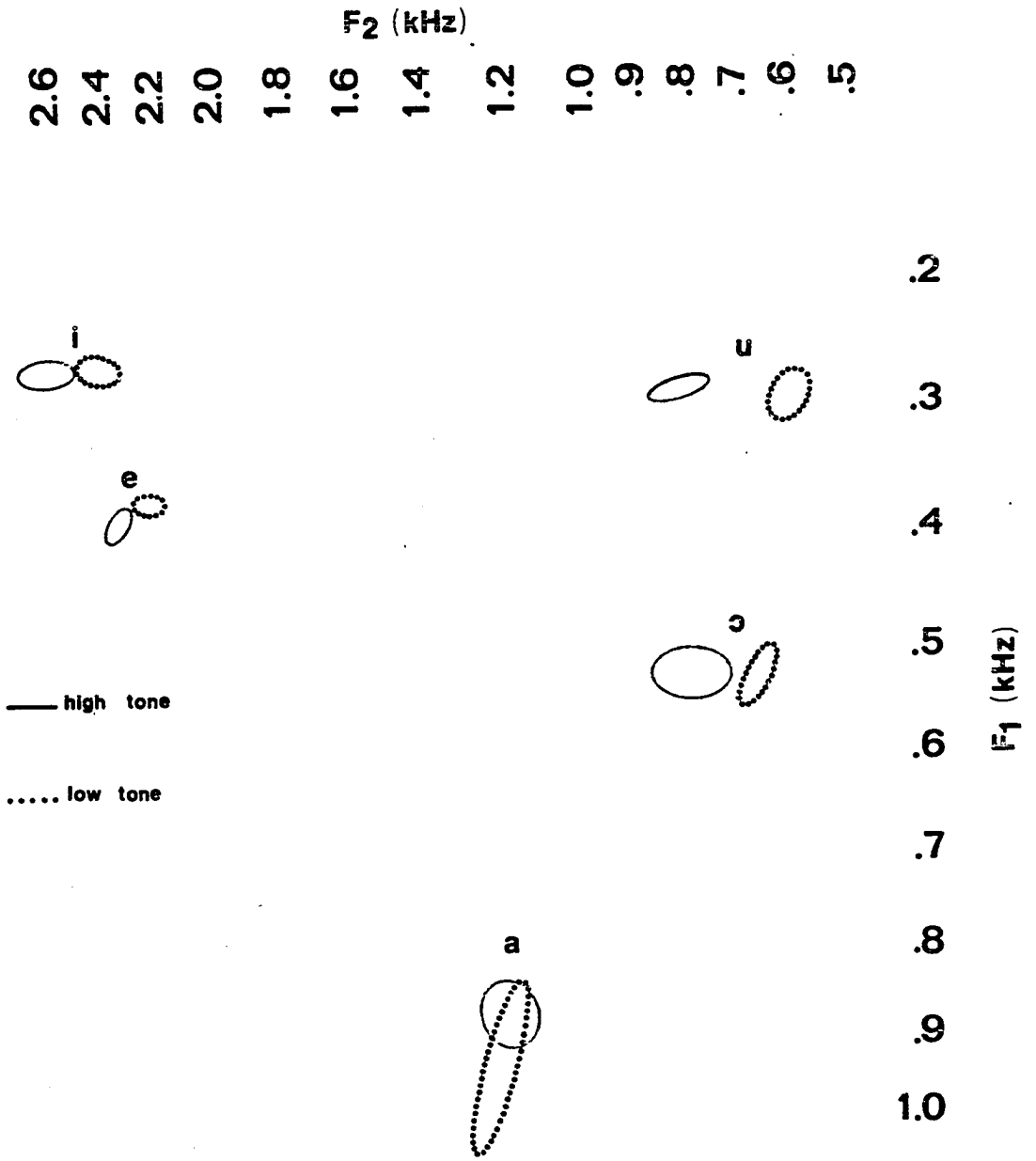


Figure 2. Vowel ellipses associated with the high and the low tones for Speaker 1

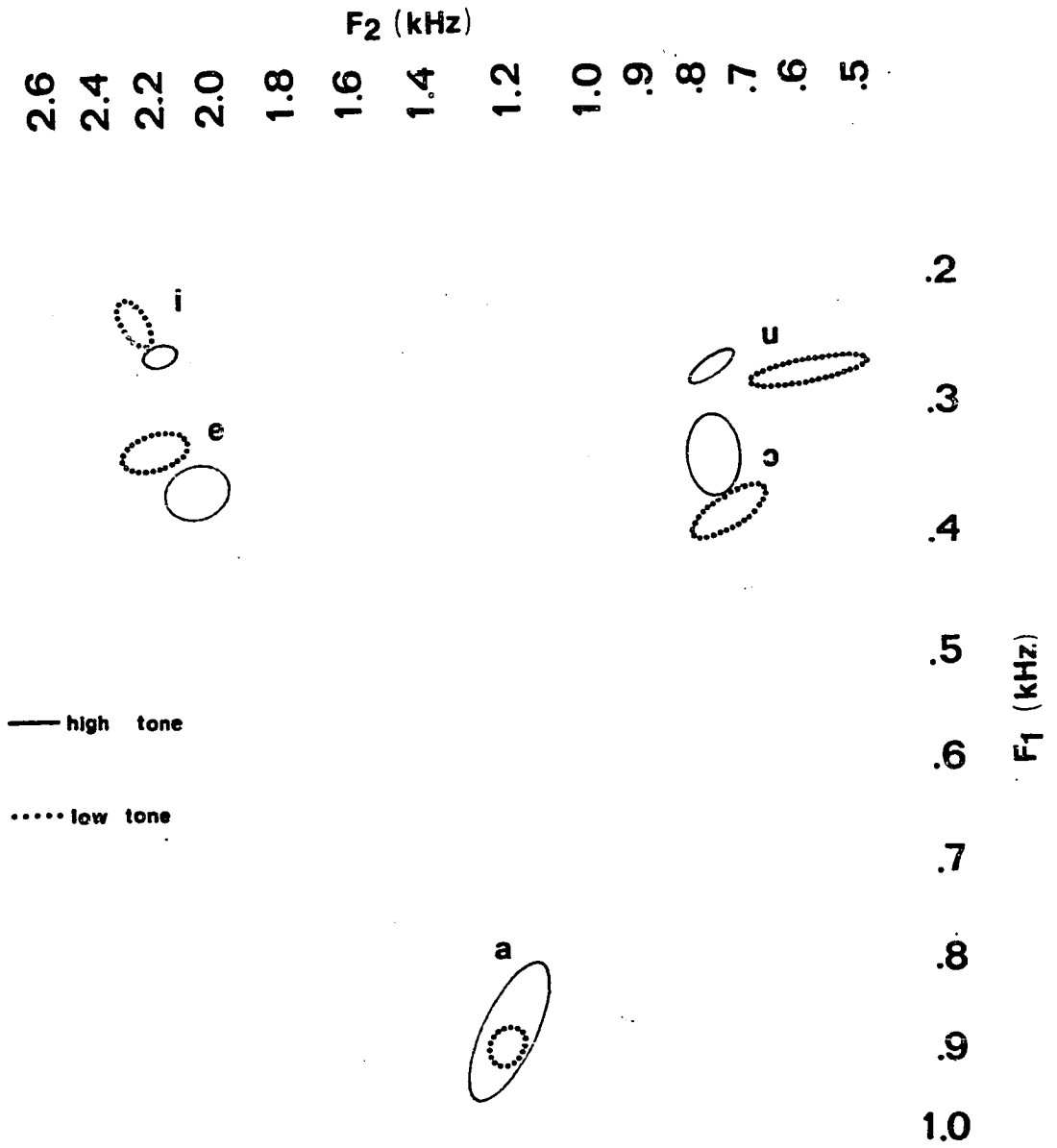


Figure 3. Vowel ellipses associated with the high and the low tones for Speaker 2

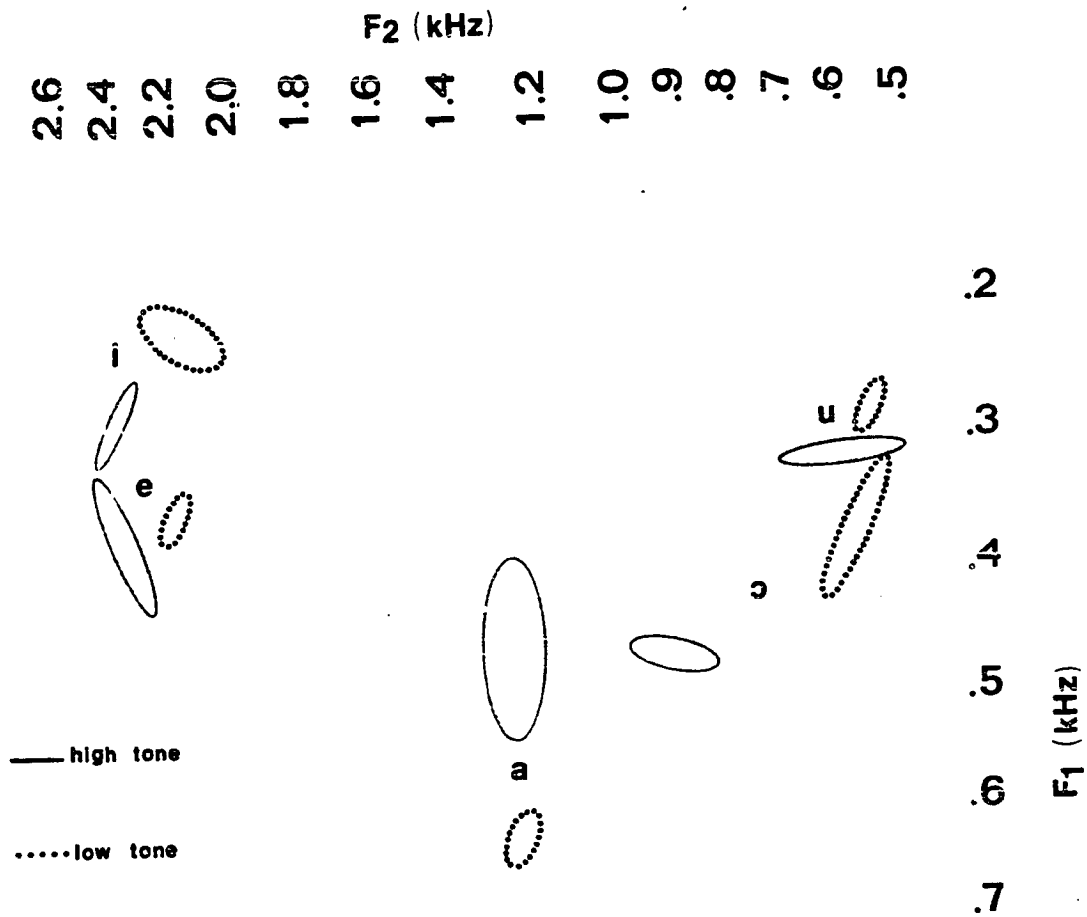


Figure 4. Vowel ellipses associated with the high and the low tones for Speaker 3

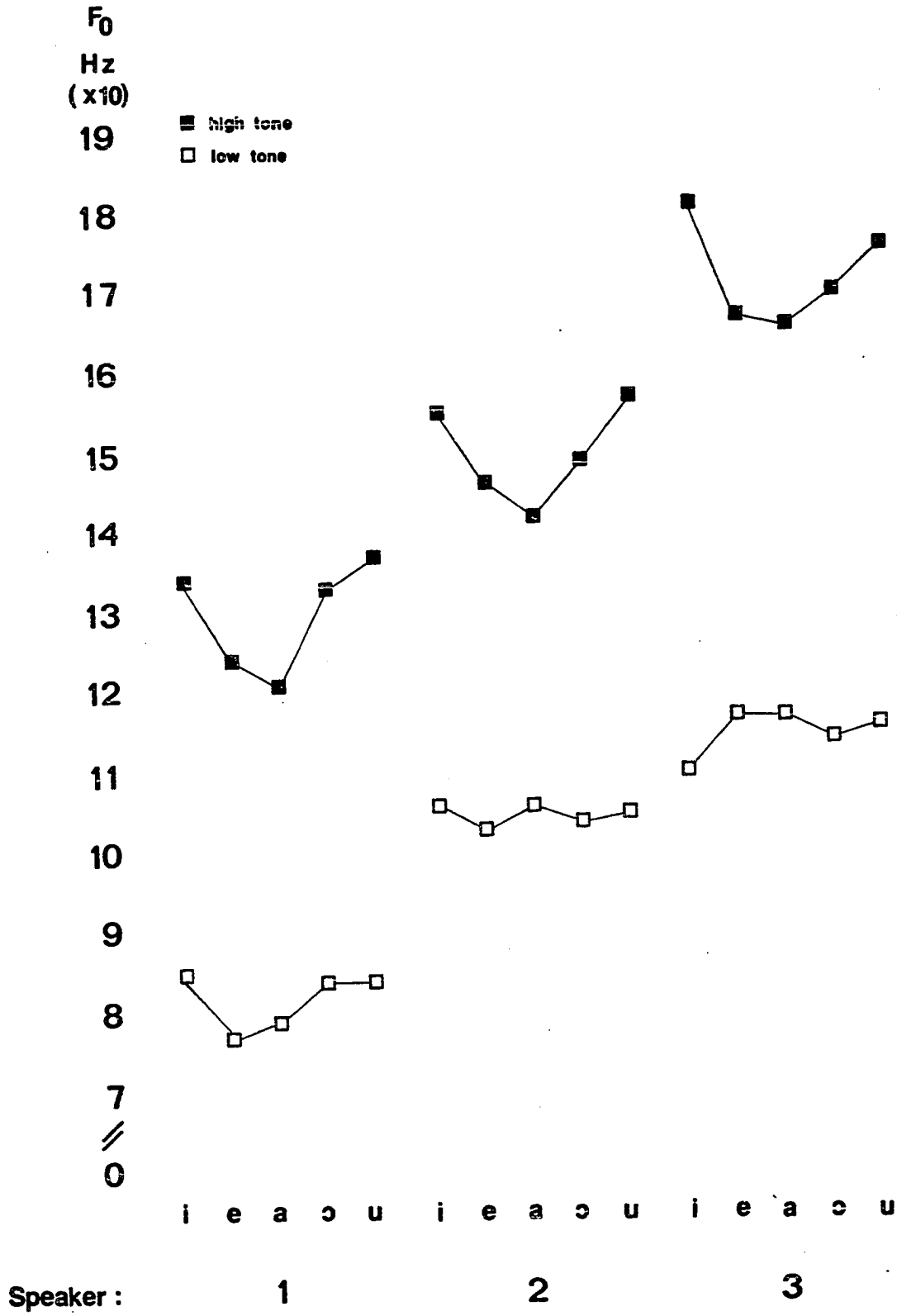


Figure 5. Average F_0 values of each tone for each vowel for Speakers 1, 2 and 3

two.

4. For all three speakers, there are no significant differences in F2 for vowel [a]. There are significant differences in F3 at the 0.01 level, but again they are not in the same direction for all three speakers.

5. There are significant differences in the same direction in F2 at the 0.001 level for vowel [u] for Speakers 1 and 2, and for vowel [ɔ] for Speakers 1 and 3, also in the same direction.

6. There are significant differences in the opposite directions in F3 at the 0.01 or 0.001 level for vowel [u] for Speakers 1 and 3.

7. These results show that vowels are in fact affected by tonal difference, although not in a systematic fashion for different speakers.

Figure 5 displays the average fundamental frequency values of the high tone and of the low tone for the five vowels produced by Speakers 1, 2 and 3. The filled squares represent the fundamental frequency values of the high tone and the open squares represent the fundamental frequency values of the low tone. Each square is the average of five tokens. The fundamental frequency value for each token is the average of the values of every 10 msec intervals. The value for each individual token is listed in Tables I, II and III. As far as the high tone is concerned, the values of the fundamental frequency correlate with vowel height. As for the low tone, the values of fundamental frequency and vowel height do not exhibit such a correlation for all three speakers.

4. DISCUSSION

It is nicely attested that the production of tone involves larynx movement, that a tone of higher fundamental frequency requires a higher larynx position (for a comprehensive review of the literature, see Ohala, 1973), and that a raised larynx shortens the length of a speaker's vocal tract from the glottis to the lips. Since the formant frequencies are to a first approximation inversely proportional to the length of a speaker's vocal tract (Fant, 1973), the acoustic consequences of different larynx height should be in general to increase or decrease all formant frequencies. Thus, by lowering the larynx in a simulated articulatory model, Lindblom and Sundberg (1971a) showed that all formant frequencies were lowered. Similarly, Riordan (1977) demonstrated that in computer simulated vowels, i, u and y, there was a consistent increase of all formant frequencies as the larynx was raised. However, our results do not show any consisten increase in the formant frequencies for vowels associated with the high tone for all three speakers. Moreover, some formant frequencies for certain vowels associated with the low tone are greater than those associated with the high tone. For instances, F1 and F3 of vowel [a], F3 of vowel [i] and F3 of vowel [ɔ] for Speaker 1; F1 of vowel [u], F2 of vowels [i] and [e] and F3 of vowel [a] for Speaker 2; and F3 of vowel [u] for Speaker 3 (See Table IV). This

seems to indicate that shifts in other articulators, besides the rise of larynx, have taken place in the production of a vowel of a higher fundamental frequency, and these shifts have contributed to the modification of the formant frequencies in addition to the modifications due to larynx movement. In fact, in an X-ray investigation, Parmenter et. al. (1933) showed that articulatory shifts accompanying an upward change in pitch of one octave involve not only the larynx alone, but also the pharynx width, the piglottis, the hyoid bone, the tongue, the jaw and the lips. These articulatory shifts seem to be the result of interaction between the anatomical effect of the larynx movement and the articulatory configuration required for the production of the vowels. Anatomically, all major articulators are connected tissues and muscles. Larynx movement will certainly affect the positions of other articulators which in turn affect the formant frequencies. Vowel perception may be another contributing factor to the modification of formant frequencies. Ainsworth (1976), in his study of fundamental frequency as the second determinant of vowel quality, reported that compensation in F1 and F2 was needed to retain phonemic identity of vowels when F_0 was submitted to large changes (from 120 Hz to 240 or 260 Hz). Although the average fundamental frequency difference between the high tone and the low tone is approximately 50 Hz for all five vowels for all three speakers, compensation in formant frequencies might be expected to occur in order to preserve the phonetic quality of the vowels. As a result, as Parmenter et. al. also pointed out, when the pitch is raised or lowered changes may be made in the size and shape of the supraglottal cavities in order to maintain the best resonance for the overtones which characterize the quality of the vowel. Our case seems to be more complicated as the low tone in Taiwanese is in fact a low falling tone (Zee, 1976). The exact way in which falling tone affects the identification of vowels is yet to be explored. It is possible to determine under laboratory conditions both the anatomical effect on the positioning of the articulators due to the larynx movement, and the perceptual effect on vowel quality due to pitch changes. However, it is much less straightforward to determine the interplay between them that eventually yields the vocal tract shape and size used by a tone language speaker producing a vowel with different tones. The formant frequency pattern associated with the high tone differs from the formant frequency pattern associated with the low tone in a way that is unique for all three speakers. This may be partly due to the differences in the auditory system and in the vocal tract shape and size among the speakers. But it may also be due to the possibility that each speaker uses different articulatory control and perceptual processing to produce a high tone or a low tone for a vowel.

5. SUMMARY OF RESULTS AND CONCLUSION

1. The formant frequencies associated with the high tone are not nece-

ssarily greater than the formant frequencies associated with the low tone for all three speakers.

2. For all three speakers, vowels [i], [e], [ɔ] and [u] associated with the high tone occupy areas in the acoustic vowel space that are distinct from those occupied by vowels associated with the low tone.

3. The way the formant frequency pattern associated with the high tone differs from the formant frequency pattern associated with the low tone is unique for all three speakers.

4. The value of the fundamental frequency of the high tone correlates with vowel height for all three speakers. No such correlation is found in vowels associated with the low tone for all three speakers.

5. Conclusion: vowels are in fact affected by tonal difference, although not in a systematic fashion for different vowels and for different speakers.

REFERENCES

Ainsworth, W. A. (1975). Intrinsic and extrinsic factors in vowel judgements. In Auditory Analysis and Perception of Speech (Fant & Tatham, eds.), N.Y., Academic Press, pp. 103-111.

Davis, S.B. (1977). Computer evaluation of laryngeal pathology based on inverse filtering of speech. SCRL Monograph #13.

Ewan, W. G. and R. Krones (1972). Laryngeal height during VCV utterances: umbrometer measurements. Paper presented at the 2nd California Linguistics Association Conference, Santa Cruz, California

Ewan, W. G. and R. Krones (1974). Measuring larynx movement using the thyroumbrometer. J. of Phonetics, 2:327-335.

Fant, G. (1973). Speech Sound and Features. Cambridge, Mass., MIT Press.

Ladefoged, P. (1967). Three Areas of Experimental Phonetics. London, Oxford Univ. Press.

Lindblom, B. E. F. and J. E. F. Sundberg (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. JASA, 50: 1166-1179.

Ohala, J. (1973). The physiology of tone. In Consonant Types and Tone (L. Hyman, ed.), L.A., University of Southern California Press.

Parmenter, C. E., S. N. Trevino and C. A. Bevans (1933). The influence of a change in pitch on the articulation of a vowel. Language, 9.1:72-81.

Riordan, C. J. (1977). Control of vocal tract length in speech. JASA, 62: 998-1002.

Sundberg, J. E. F. and P.-E. Nordström (1976). Raised and lowered larynx - the effect on vowel formant frequencies. STL-QRSP, 2-3: 35-39.

Vanderslice, R. (1967). Larynx vs. lungs: cricothyrometer data refuting some recent claims concerning intonation and "archetypality". UCLA WPP #7: 69-79.

Zee, E. (1976). Duration and intensity as correlates of F_0. Paper presented at the 92nd ASA meeting, Autumn, 1976, San Diego, California.

Section 3 *The Effect of Aspiration on the F_0 of the Following Vowel in Cantonese*

1. INTRODUCTION

Conflicting results have been reported with regard to the effect of voiceless aspirated stop consonants on the F_0 onset of the following vowel. Han and Weitzman (1967, 1970) demonstrated that for their Korean subjects the F_0 onsets of vowels following the aspirated stops [p^h, t^h, k^h] are much higher than the F_0 onsets of vowels following the weak aspirated stops [p, t, k], although the difference between the F_0 onsets of vowels after the aspirated stops and the strong unaspirated stops [P̄, T̄, K̄] is much smaller. A separate study of the Korean stops (Kagaya, 1974) produced inconclusive results with respect to the effect of the aspirated stops on the F_0 onset of the following vowel as the two subjects who participated in the experiment produced conflicting results. Jeel (1975) reported that in the speech of six Danish speakers the F_0 onset of a vowel after the asiprated stops [p^h, t^h, k^h] is consistently higher than that of a vowel following the unaspirated stops [p, t, k]. In Erickson (1975), eight of the eleven Thai subjects had a higher F_0 onset for a vowel following the aspirated stop [p^h] than for a vowel following the unaspirated stop, whereas the other three produced opposite results. Ewan (1976) reported that in the speech of a Japanese and a Thai speaker the F_0 onset of a vowel after a voiceless aspirated stop [p^h] is also higher. However, in the analysis of the speech of a Thai speaker, Gandour (1974) presented data showing that the F_0 onset of a vowel after the voiceless aspirated stops [p^h, t^h] is slightly lower than the F_0 onset of a vowel after the unaspirated counterparts. Kagaya and Hirose (1975) showed that in the speech of a Hindi speaker the F_0 onset of a vowel following an aspirated stop is also slightly lower. Hombert & Ladefoged (1977) concluded that voiceless aspirated and voiceless unaspirated stops have similar effects on the F_0 of the following vowel. It seems no agreement can be reached with respect to the effect of the voiceless aspirated stops on the F_0 onset of the following vowel. The disagreement certainly requires further research in this aspect. In the present study we investigate the difference between the effect of [p^h] and [p] on the F_0 onset of the following diphthong [ei] in Cantonese. In this Chinese dialect, the high tone, historically the Ying-Ping tone, may occur on the vowel in both syllable types [p^hV] and [pV]. Etymologically, neither [p^h] nor [p] were derived from [b^h], so a difference in F_0 onset of the following diphthong between the syllable types cannot be attributed to an earlier voicing difference.

2. PROCEDURE

A reading list was prepared containing 10 repetitions of two Cantonese words, [p^hei] 'to spread' and [pei] 'sorrow' in the sentence frame below:

[ŋo	yiu	tvk	___	pei	nei	t^hisn]
I	want	read		for	you	listen

Other dummy meaningful words were added to the reading list in order to avoid monotony which may be caused by the limited number of the test words. The tokens in the reading list were arranged in a random order. Three male

native Cantonese speakers participated in the investigation. They were undergraduate students in their early twenties. Each speaker was asked to read the word list at a normal rate of speech. The recording was performed in a single session for each speaker in a sound treated booth. The recorded tapes were analyzed using a PDP-12 computer. A fundamental frequency measurement for each test word was obtained every 10 msec by the Cepstrum method with a window size of 51.2 msec. Also obtained every 10 msec were the intensity (rms) values of the test words, using a square window of 51.2 msec.

3. RESULTS

The F_0 contours of the vocalic portion of all the tokens are shown in Figure 1 for the speech of the three speakers. Each dot represents one data point of the output of the Cepstrum analysis and the time interval between any two successive dots is 10 msec. On the left of the figure are the F_0 contours for the 10 tokens associated with the aspirated stop [p^h] and on the right are the F_0 contours for the 10 tokens associated with the unaspirated stop [p]. It can be easily seen that for all three speakers the values of the initial data points are greater for the F_0 contours associated with [p^h] than those associated with [p].

Table I shows the values of the F_0 onsets and the intensity (rms) onsets for the tokens associated with [p^h] or [p] for all three speakers. The F_0 onset is defined as the mean value of the first three data points of the output of the Cepstrum analysis. Similarly, the intensity onset is the mean value of the first three data points of the output of the intensity measurement. These numbers thus reflect values in the first 71.2 msec following the consonant. Also shown in the table are the means ($\bar{X}$) and the standard deviations (S.D.) of the F_0 and intensity onsets for each set of 10 tokens associated with [p^h] or [p]. In the bottom of the table are the results (t-scores) of the grouped data t-tests (one tailed) between the F_0 onsets associated with [p^h] and [p], and between the intensity onsets associated with [p^h] and [p]. These results show that the difference between the F_0 onsets associated with [p^h] and those associated with [p] are highly significant at the 0.005 level for all three speakers. This is also true for the difference between the intensity onsets associated with the two types of stop consonants.

The values of the F_0 and the intensity onsets for all the tokens by all three speakers shown in Table I are plotted in Figure 2. The circles and the filled circles represent the F_0 onsets associated with [p^h] and [p] respectively, and the empty squares and the filled squares represent the intensity onsets associated with [p^h] and [p] respectively. The circle and the empty square, or the filled circle and the filled square in the same column refer to the F_0 and the intensity onsets of a single token. We can see that for all the tokens the F_0 onsets associated with [p^h] are higher than those associated with [p] for all three speakers. However, the corres-

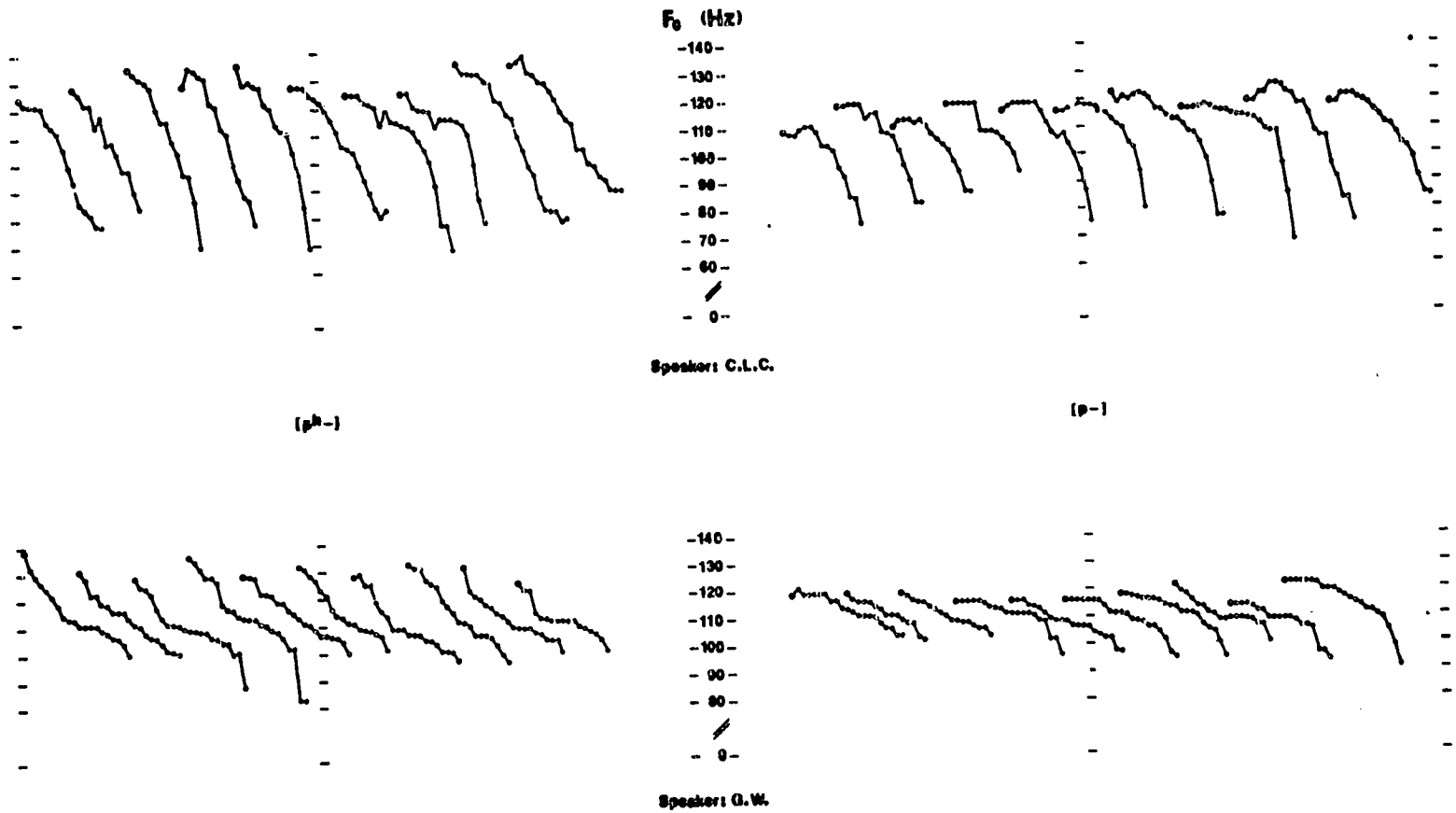


Figure 1. F_0 contours of the diphthong [eɪ] following [pʰ] (on the left) and [p] (on the right) by speakers C.L.C. and G.W.

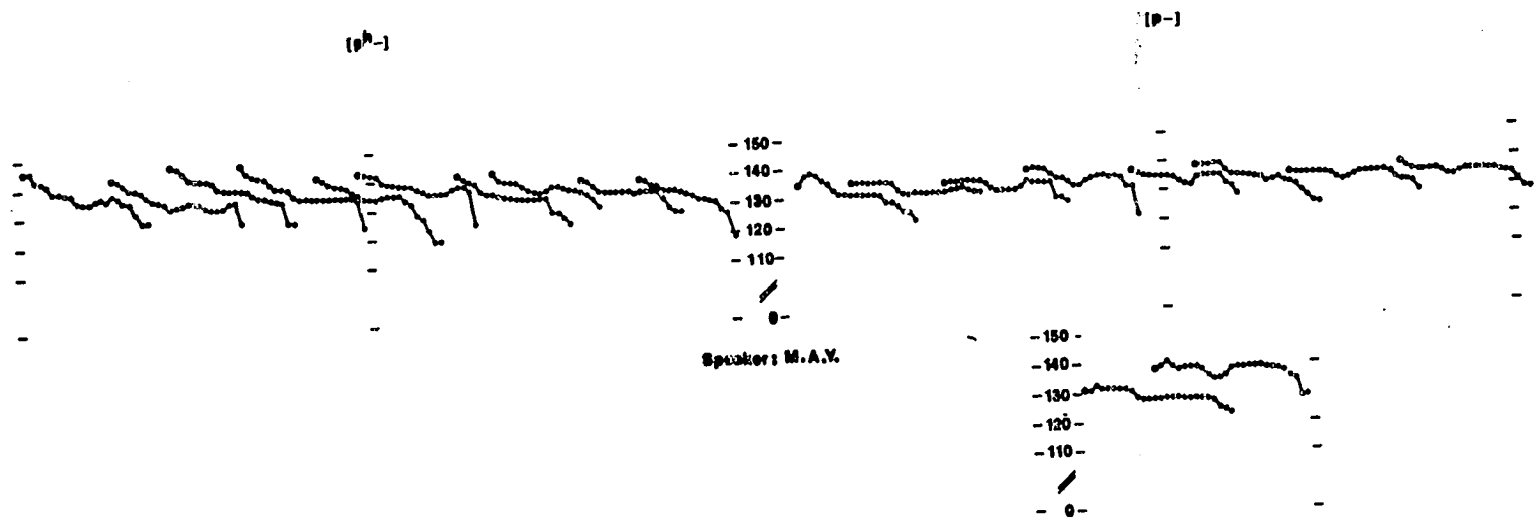


Figure 1. F_0 contours of the diphthong [eɪ] following [pʰ] (on the left) and [p] (on the right) by speaker M.A.Y.

	[pʰ]		[p]		[pʰ]		[p]		[pʰ]		[p]	
	rms	F0	rms	F0	rms	F0	rms	F0	rms	F0	rms	F0
	1284	127.4	1969	116.2	881	143.2	1268	139.9	1266	122.8	1436	119.4
	1245	131.9	1619	116.2	992	142.3	1011	136.6	1331	125.5	1588	108.8
	1227	124.4	2106	122.2	955	146.3	1183	135.2	1395	133.1	1483	112.7
	1228	127.2	1754	116.4	1020	145.1	1123	138.7	1304	132.6	1460	119.1
	1021	126.0	1731	118.0	994	144.7	1036	131.9	1364	131.1	1394	116.4
	1182	134.1	1428	113.6	1073	141.1	1101	135.1	1330	128.0	1419	121.1
	1125	128.4	1690	119.5	1061	138.2	1158	135.0	1347	125.0	1557	116.2
	1021	130.2	1672	120.0	964	140.1	1142	138.8	1429	123.6	1598	118.0
	910	123.0	1697	117.9	1042	140.0	1120	136.3	1378	132.8	1421	118.8
	1229	131.8	1698	117.2	950	138.8	1121	136.7	1325	136.4	1433	120.1
$\bar{X}$:	1147	128.4	1738	117.7	993	142.0	1126	136.4	1347	129.1	1497	117.1
S.D.:	124	3.5	185	2.4	59	2.6	72	2.3	47	4.8	77	3.8
Speaker:	G.W.				M.Y.				C.L.C.			

	F0		rms		F0		rms		F0		rms	
	[pʰ]	[p]	[pʰ]	[p]	[pʰ]	[p]	[pʰ]	[p]	[pʰ]	[p]	[pʰ]	[p]
t-score:	7.9		8.4		4.8		4.5		6.2		4.7	
p:	<.005		<.005		<.005		<.005		<.005		<.005	

Table I. Values of the F0 onsets and the intensity onsets for the tokens associated with [pʰ] or [p] for all three speakers, and results (t-scores) of the grouped data t-tests (one tailed) between the F0 onsets associated with [pʰ] and [p] and between the intensity onsets assoicated with [pʰ] and [p].

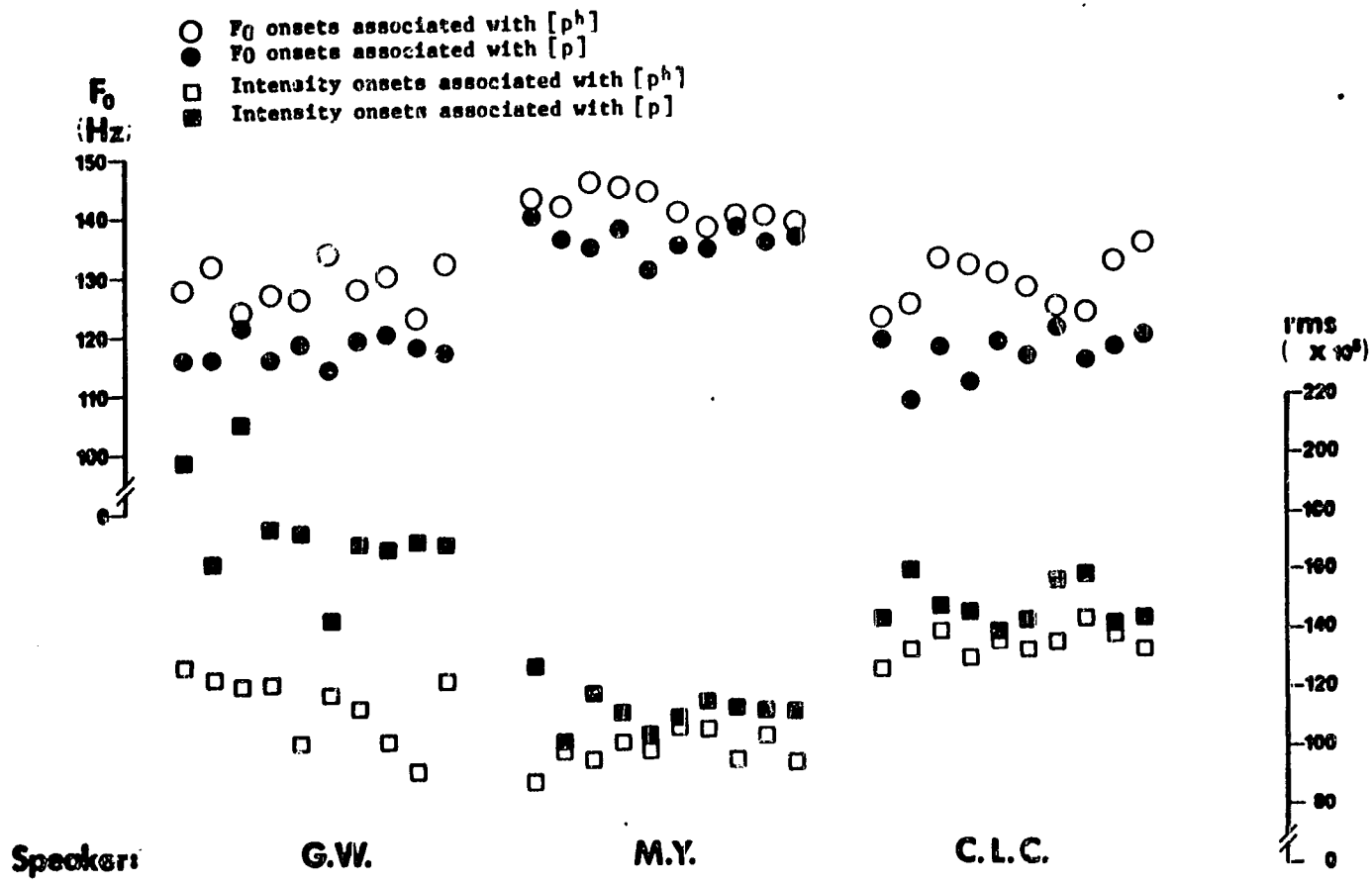


Figure 2. Plotted values of the F0 and the intensity onsets for all tokens by all three speakers.

ponding intensity onsets associated with [p^h] are lower than those associated with [p], and this is true for all three speakers. Thus, we have shown that in the speech of three Cantonese speakers the high tone on the diphthong [ei] after [p^h] has a higher F_0 onset but a lower intensity onset than the same tone on the same diphthong following [p].

4. DISCUSSION.

The effect of the voiceless aspirated stops on the F_0 onset of the following vowel not only varies according to different languages, but also differs according to individual speaker within a language, for instance, both in Kagaya (1974) on Korean and Erickson (1975) on Thai, speakers of the same language have produced different results. Our findings in Cantonese have not contributed to resolve the issue, nevertheless, the Cantonese data has increased the number of the languages in which voiceless aspirated stops raise the F_0 onset of the following vowel. That the intensity onset of the diphthong [ei] following [p^h] is always lower in Cantonese seems to imply that the subglottal pressure at the onset of the diphthong is also lower. The fact that the F_0 onset of the diphthong following [p^h] is always higher indicates that a higher F_0 may be produced even with a decreased subglottal pressure. However, it is not clear at this point why opposite results are produced by speakers of different languages, or by speakers of the same language. In order to have a better understanding of the causes of such differences, future investigations should obtain data on airflow, subglottal pressure, larynx height, glottal aperture, and vocal cord lengths at the onset of a vowel following [p^h] or [p].

REFERENCES

Erickson, D. (1975). Phonetic implications for a historical account of tonogenesis in Thai. In J.G. Harris & J.R. Chamberlain (eds.), Studies in Tai Linguistics in Honor of W.J. Gedney. Bangkok, Central Institute of English Language, Office of State Universities, 100-111.

Ewan, W.G. (1976). Laryngeal behavior in speech. University of California, Berkeley Ph.D. Dissertation.

Gandour, J.T. (1974). Consonant types and tone in Siamese. Journal of Phonetics 2, 337-350.

Han, M.S. & Weitzman, R.S. (1967). Acoustic features in the manner-differentiation of Korean stop consonants. Studies in the Phonology of Asian Languages, V. Los Angeles, Acoustic Phonetics Research Lab, University of Southern California.

Han, M.S. & Weitzman, R.S. (1970). Acoustic features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/. Phonetica 22, 112-128.

Hombert, J.-M. & Ladefoged, P. (1977). The effect of aspiration on the fundamental frequency of the following vowel. UCLA Working Papers in Phonetics 36, 33-40.

Jeel, V. (1975). An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. Annual Report of the Institute of Phonetics, University of Copenhagen 9, 191-211.

Kagaya, R. (1974). A fiberoptic and acoustic study of the Korean stops, affricates and fricatives. Journal of PHonetics 2, 161-180.

Kagaya, R. & Hirose, H. (1975). Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. Annual Bulletin of Research Institute of Logopedics and PHoniatrics 9, 27-46.

Section 4 *The Effect of F_0 on the Duration of* [s]

The effect of F_o on the duration of [s]

1. Introduction

There have been studies concerning the influence of vowels on the duration of an adjacent fricative /s/. Klatt (1974) reported that in English there is a tendency for the duration of prestressed /s/ to be longer than the duration of preunstressed /s/. Schwartz (1970) found that the fricative /s/ in the environment of a high front vowel /i/ is longer than in the environment of a low vowel /a/. The purpose of the present study was to determine whether the duration of the prevocalic fricative /s/ is influenced by the tone on the following vowel. It also examines whether temporal compensation exists between the duration of the fricative and the duration of the vowel.[1] A tone language, Taiwanese, was used for investigation. In this Chinese dialect, all its five contrastive tones; high, mid, low, low-rising and high-falling, occur on the syllable /si/:

/sí/	(high)	meaning	'silk'
/sī/	(mid)	"	'yes'
/sì/	(low)	"	'four'
/sǐ/	(low-rising)	"	'time'
/sî/	(high-falling)	"	'to die'

2. Procedure

A word list containing 75 tokens of these syllables was prepared, that is, 15 repetitions for each tone class. Two male Taiwanese speakers, both in their early thirties, were instructed to read the list at a normal rate of speech. The test word was placed in a carrier frame as follows:

guá tçí-má t^hà? ___ taì-kēi t^hía̍

'I now read ___ everyone listens'

The recording was made in a sound treated room in a single session for each speaker. The data were analyzed, using a PDP-12 computer at the UCLA Phonetics Laboratory. Duration measurements of both segments of the syllable /si/ were made directly on the waveform.

3. Results

Figure 1 shows the results of the duration measurements of both the fricative and the vowel of the syllable produced with different tones for both speakers. Each blank bar on the left represents the mean duration of 15 tokens of the fricative /s/ for each tone class. Each dark bar on the right represents the mean duration of 15 tokens of the vowel /i/ for each tone class. Durations are given in milliseconds..

An examination of the durations of /s/ for both speakers shows that the duration of /s/ associated with either the low tone or the low-rising tone is longer than the duration of /s/ associated with the high, the mid, or the high-falling tone.

T-test results reveal that the difference in duration between /s/ associated with the low tone and /s/ associated with the high, the mid, or the high-falling tone is significant.[2] The difference in duration between /s/ associated with the low-rising tone and /s/ associated with the high, the mid, or the high-falling tone is also significant.[3]

The fact that the duration of the fricative /s/ associated with either the low or the low-rising tone is longer than the fricative /s/ associated the high, the mid, or the high-falling tone indicates that there may be a correlation between the duration of the fricative /s/ and the onset value of the tone on the following vowel /i/.

4. Disucssion

In Zee and Hombert (1976), we presented the duration pattern of the vocalic portion of the syllable /si/ produced with the same five Taiwanese tones, which is similar to the duration pattern obtained in the present study as shown on the right side of Figure 1. The low-rising tone is longer than the high and the mid tones, and the high and the mid tones are longer than the low and the high-falling tones. It was claimed that the duration of the vocalic portion of the syllable /si/ is determined by the shape of the tone. Figure 2 shows the duration of the entire syllable produced with five different tones. Each is the mean of 15 tokens. The only difference between Figure 1 and Figure 2 is that in Figure 1 the lineup point is the onset of a vowel and in Figure 2 is the beginning of the fricative /s/. As before, the blank bars represent the fricative portion /s/ and the dark bars represent the vocalic portion /i/. We can see for both speakers the duration of the syllable /si/ associated with the low-rising tone is longer than the duration of the syllable associated with any other tones; and the duration of the syllable associated with the high-falling tone is the shortest of all. Although the duration of the vocalic portion associated with the low tone is similar to the duration of the vocalic portion associated with the high-falling tone, the duration of the syllable is longer for the low tone than for the high-falling tone as the duration of the fricative portion for the low tone is longer than the high-falling tone. Thus, the duration of a syllable, such as /si/, varies according to the tone that the vowel carries, and furthermore, it is the onset of the tone

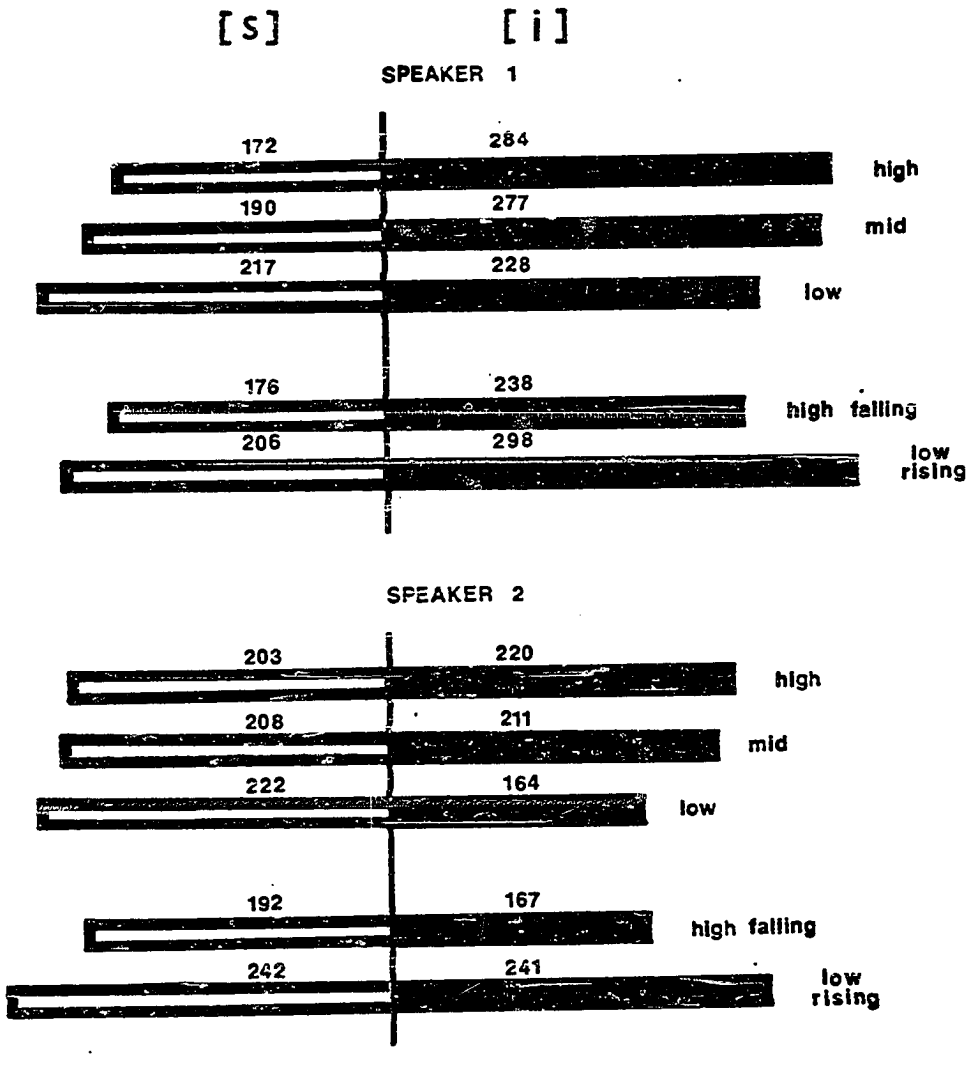


Fig. 1

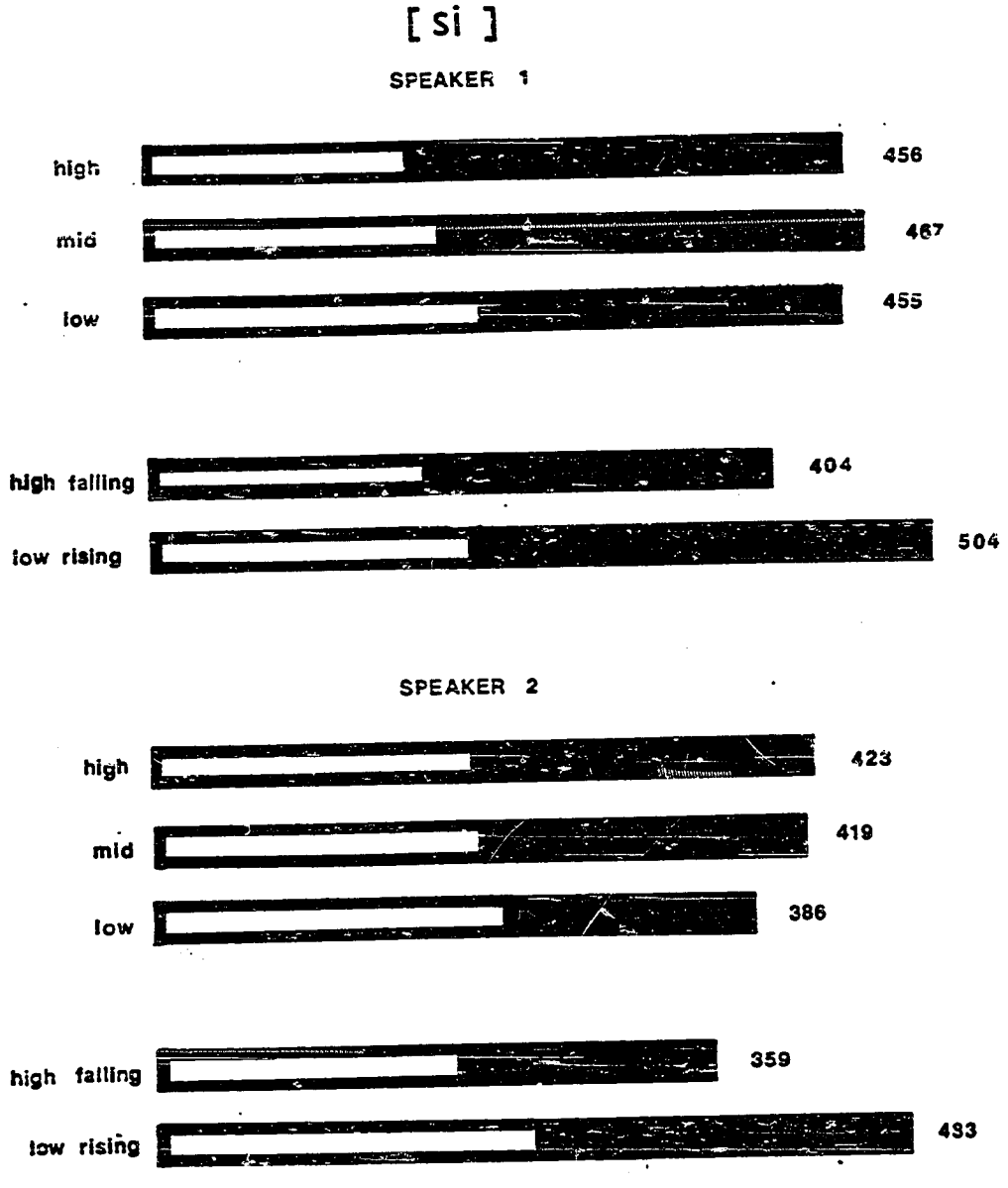


Fig. 2

that determines the duration of the fricative portion and the shape of the tone that determines the duration of the vocalic portion.

Temporal compensation within the sequences of the segments /s/ and /i/ associated with any one of the five tone classes was tested. To see whether there is temporal compensation within two segments is to find out whether the durations of the two segments are negatively correlated (Lehiste, 1970). There is a negative correlation between the durations of two segments if the relative variance of the duration of the sequence of two segments is less than the sum of the relative variances of the segments considered separately. Relative variance is simply variance divided by average duration, a concept introduced by Allen (1969). The result is, within each tone class, the relative variance of the duration of the syllable /si/ as one unit is greater than the sum of the relative variances of the segments /s/ and /i/ considered separately. This is true for both speakers. This indicates that there is not a negative correlation between the durations of the two segments. It follows that, within each tone class, temporal compensation does not exist between the duration of the segment /s/ and the duration of the segment /i/ in the syllable /si/.

A negative correlation between the durations of two segments has been taken to imply that the two are programmed as one unit at some higher level at which articulatory sequences are programmed (Lehiste, 1970). Our result would seem to imply that the segments /s/ and /i/ in the syllable /si/ produced with any tone are programmed not as one unit but as two separate units at some higher level, as there is not a negative correlation between the durations of the two segments /s/ and /i/. However, the fact that the duration of the segment /s/ is determined by the tone on the following vowel /i/ indicates that the segments /s/ and /i/ in the syllable /si/ produced with any tone cannot be programmed as two separate units at some higher level at which articulatory sequences are programmed, but as one unit. Thus, in a tone language, such as Taiwanese a negative correlation between the durations of two segments, a fricative /s/ followed by a vowel /i/, does not constitute a necessary condition for determining whether the two segments are programmed as two separate units, or as one unit, at some higher level at which articulatory sequences are programmed.

5. Summary and Conclusions

The purpose of this study was to determine whether the duration of the prevocalic fricative /s/ is influenced by the tone on the following vowel and to examine whether temporal compensation exists between the duration of the fricative and the duration of the vowel within a certain tone class. A tone language, Taiwanese, was used for investigation. The results from analyzing the data produced by two male Taiwanese speakers indicate that:

(1) the duration of the fricative /s/ is influenced by the tone on the following vowel, and it varies according to the onset value of the tone.

(2) within each tone class, temporal compensation does not seem to exist between the durations of the segments /s/ and /i/ in the syllable /si/.

(3) in a tone language, such as Taiwanese, a negative correlation between the durations of two segments, such as a fricative /s/ followed by a vowel /i/, does not constitute a necessary condition for determining whether the two segments areprogrammed as two separate units, or as one unit, at some higher level at which articulatory sequences are programmed.

Footnotes

1. It refers to the temporal compensation between segments /s/ and /i/ within one single tone class, not across different tone classes.

2. (a) The difference in duration between /s/ associated with the low tone and /s/ associated with the high tone is significant at the .005 level for Speaker 1 and .025 level for Speaker 2.

(b) The difference in duration between /s/ associated with the low tone and /s/ associated with the mid tone is significant at the .005 level for Speaker 1. However, only a trend in this direction was obtained for Speaker 2 (.1 level).

(c) The difference in duration between /s/ associated with the low tone /s/ associated with the high-falling tone is significant at the .005 level for both Speaker 1 and 2.

3. The difference in duration between /s/ associated with the low-rising tone and /s/ associated with the high, the mid, or the high-falling tone is significant at the .005 level for both speakers.

References

Allen, George D. (1969) Structure of timing in speech production. Paper presented at the 78th Acoust. Soc. of Amer. Meet. Autumn, 1969, SanDiego, Calif.

Klatt, Dennis (1974) The duration of [s] in English words. *Journ. of Speech and Hearing Res.* 17, 51-63.

Lehiste, Ilse (1970) Temporal organization of spoken language. The Ohio State University, *Working Papers in Linguistics*, 4, 96-114.

Schwartz, Martin F. (1969) Influence of vowel environment upon the duration of /s/ and /š/. *Journ. of the Acoust. Soc. of Amer.* 46, 480 (L).

Zee, Eric and Jean-Marie Hombert (1976) Duration and intensity as correlates of F_o. Paper presented at the 92nd ASA meeting, Autumn, 1976, San Diego, Calif.

Section 5 *Peak Intraoral Air Pressure in [p] as a Function of F_0*

1. INTRODUCTION

There have been a number of studies investigating the differences in intraoral air pressure between voiced and voiceless stop consonants. The general agreement among investigators is that voiceless stop consonants are produced with higher peak intraoral air pressure than their voiced cognates (Black, 1950, Malecot, 1955, 1966, Warren, 1964, Subtelny, Worth and Sakuda, 1966, Arkebauer, et al, 1967, Ringel, House and Montgomery, 1967, Löfqvist, 1971, Brown, McGlone and Profitt, 1973, Bernthal, 1978). The differences in pressures have usually been attributed to the effect of glottal resistance (Malecot, 1955, Warren, 1964, Arkebauer, et al, 1967). During the production of voiced stop consonants, the adduction of the vocal cords increases the resistance to pulmonic air entering the vocal tract and thus reduces the pressures in the oral cavity. With regard to the consonantal position in the syllable, it has been shown that for voiceless stop consonants the peak intraoral air pressure is higher in medial position , lower in initial position and lowest in final position (Malecot, 1955, 1968, Arkebauer, et al, 1967, Lisker, 1970, Löfqvist, 1971). The relation between tone and peak intraoral air pressure, however, has not been explored. It is the purpose of this study to investigate whether the peak intraoral air pressure in a voiceless unaspirated stop [p] is affected by the tone on the following vowel. As variations in fundamental frequency may be accompanied by changes in subglottal pressure (Ladefoged, 1963, Öhman & Lindqvist, 1966) and as the subglottal pressure and the intraoral air pressure are more or less the same during the production of a voiceless stop consonant (Ladefoged, 1967, 1968, Netsell, 1969, Scully, 1969), it is reasonable to expect that the peak intraoral air pressure in [p] may vary according to changes in subglottal pressure for the production of different tones on the vowel that follows [p].

2. PROCEDURE

In our investigation, four Chinese dialects, Standard Mandarin, Chungking (Western Mandarin), Cantonese and Shanghai, were used. In each dialect, different tones may occur on a CV syllable, where C is a voiceless unaspirated bilabial stop [p] to form words of different meaning, as shown in the following:

Mandarin			Chungking		
44	[pi]	'to oppress'	45	[pi]	'sex organ'
35	[pi]	'nose'	21	[pi]	'to oppress'
214	[pi]	'to complete'	41	[pi]	'to compare'
51	[pi]	'closed'	13	[pi]	'to avoid'

Cantonese			Shanghai		
55	[pei]	'sorrow'	53	[pi]	'edge'
33	[pei]	'secret'	34	[pi]	'to change'
22	[pei]	'nose'			

The numerals to left of each word denote levels of pitch, '5' being the highest and '1' being the lowest. Thus a combination such as '51' means a high-low falling tone, whereas '13' is a low-mid rising tone. A reading list for each dialect was prepared containing 10 repetitions of each word in the sentence frames below:

Mandarin:	[uo	cian-tsai	tu	___	kei	ni	t^hiŋ]
	I	now	read		for	you	listen
Chungking:	[ŋo	cian-tsai	tu	___	tcɩ	ni	t^hiŋ]
	I	now	read		for	you	listen
Cantonese:	[ŋo	yiu	tvk	___	pei	nei	t^hian]
	I	want	read		for	you	listen
Shanghai:	[ŋu	yɩo	do?	___	pə?	i	t^hiŋ]
	I	want	read		for	him	listen

Four male native speakers, one for each dialect, all in their late thirties, participated in the experiment. Each speaker was asked to read the sentences at a normal rate of speech. He was also asked not to emphasize the test word.

During the readings intraoral air pressure was sensed with a polyethylene tube (120 mm in length and 2.5 mm internal diameter). The tube was inserted into the oral cavity between the upper and lower lips and was maintained near the center line in the oral cavity and parallel to oral airflow. The length of the tube inside the month was approximately 40 mm. Outside the mouth, the tube was attached to a pressure transducer. Transduced intraoral air pressure signals were amplified and displayed on an oscilloscope. To obtain measureable records the oscilloscope was in turn connected to one channel of an oscillomink and a microphone was connected to another. The intraoral air pressure recording system was calibrated in cm H_2O against a U-tube water manometer. The accuracy of the recording was calculated to be within the range of ± 0.24 cm H_2O. The recording of the intraoral air pressure and the utterances was performed in a sound treated booth. The audio signals were analyzed using a PDP-12 computer. Fundamental frequency measurements for each test word were obtained every 10 msec by the Cepstrum method, using a window of 51.2 msec.

3. RESULTS

Table Ia presents the values (in cm H_2O) of the peak intraoral air pressure for [p] associated with different tones on the following vowel or diphthong for each of the four speakers. Also shown are the means and the standard deviations of these values for [p] associated with each tonal category. The mean of the F_0 onset values for the vowel or the diphthong associated with each tonal category is given at the bottom of the table. The F_0 onset refers to the average of the first three data points of the output of the Cepstrum analysis. We can see in general a higher mean peak intraoral air pressure is associated with a higher mean F_0 onset, with

SPEAKER	Mandarin				Chungking				Cantonese			Shanghai	
TONE TYPE	44	35	214	51	45	21	51	13	55	33	22	53	34
PEAK INTRA-ORAL AIR PRESSURES (cm H_2O)	4.36	3.76	3.80	4.52	3.76	4.00	4.76	4.24	4.60	4.56	4.24	3.56	3.12
	4.24	3.88	3.88	5.36	4.76	3.60	4.52	3.36	4.64	4.36	4.12	3.56	2.52
	4.44	3.24	3.76	5.24	4.52	3.76	4.00	3.60	4.56	4.20	4.12	4.12	3.24
	3.76	3.52	3.12	4.60	3.49	4.52	4.12	3.60	4.64	4.44	4.28	3.80	3.00
	4.60	3.64	3.36	4.24	3.24	3.76	4.24	3.60	4.80	4.32	4.12	3.64	3.00
	4.00	3.68	3.52	5.32	4.64	3.00	5.00	3.24	4.80	4.36	4.00	3.92	3.32
	4.56	3.52	4.00	4.44	4.12	3.12	3.88	3.40	4.52	4.24	3.76	4.92	3.44
	4.12	3.24	4.00	4.48	3.64	4.76	4.12	3.60	4.52	4.52	4.36	3.20	3.32
	4.10	3.52	2.64	4.64	4.00	4.24	4.24	--	4.52	4.00	3.88	4.12	3.52
	4.60	--*	--	--	4.24	--	4.88	--	4.36	4.64	3.88	4.12	3.36
$\bar{X}$	4.28	3.56	3.56	4.76	4.04	3.86	4.38	3.58	4.60	4.36	4.08	3.82	3.18
S.D.	0.28	0.22	0.46	0.43	0.51	0.60	0.39	0.29	0.13	0.19	0.19	0.32	0.29
F_0 ONSET (Hz) $\bar{X}$	216	140	99	248	150	112	171	99	188	150	137	166	126
F_0 ONSET (Hz) S.D.	8.2	3.0	5.4	10.6	4.7	2.6	5.6	2.7	6.3	4.5	5.0	15.4	5.7

* '--' token mispronounced.

Table Ia. Values (cm H_2O) of the peak intraoral air pressure for [p] associated with different tones on the following vowel or diphthong, and mean F_0 onsets (Hz) of the vowel or diphthong for each of the four speakers.

SPEAKER	TONE TYPES & MEAN F0 ONSET	MEAN F0 ONSET DIFFERENCE	PEAK INTRAORAL AIR PRESSURES MEAN	t-SCORE	SIGNIFICANCE
Mandarin	44 (216 Hz) 35 (140 Hz)	76 Hz	4.28 3.56	6.27	p <0.005
	44 (216 Hz) 214(99 Hz)	117 Hz	4.28 3.56	4.14	p <0.005
	44 (216 Hz) 51 (248 Hz)	32 Hz	4.28 4.76	2.93	p <0.005
	35 (140 Hz) 214(99 Hz)	41 Hz	3.56 3.56	0.53	Non-sig.
	35 (140 Hz) 51 (248 Hz)	108 Hz	3.56 4.76	7.56	p <0.005
	214(99 Hz) 51 (248 Hz)	149 Hz	3.56 4.76	5.74	p <0.005
Chungking	45 (150 Hz) 21 (112 Hz)	38 Hz	4.07 3.86	0.85	Non-sig.
	45 (150 Hz) 51 (171 Hz)	21 Hz	4.07 4.38	1.64	Non-sig.
	45 (150 Hz) 13 (99 Hz)	51 Hz	4.07 3.58	2.49	p <0.025
	21 (112 Hz) 51 (171 Hz)	59 Hz	3.86 4.38	2.38	p <0.025
	21 (112 Hz) 13 (99 Hz)	13 Hz	3.86 3.58	1.17	Non-sig.
	51 (171 Hz) 13 (99 Hz)	72 Hz	4.38 3.58	5.09	p <0.005
Cantonese	55 (188 Hz) 33 (150 Hz)	38 Hz	4.60 4.36	3.17	p <0.005
	55 (188 Hz) 22 (137 Hz)	51 Hz	4.60 4.08	6.98	p <0.005
	33 (150 Hz) 22 (137 Hz)	13 Hz	4.36 4.08	3.27	p <0.005
Shanghai	53 (166 Hz) 34 (126 Hz)	40 Hz	3.82 3.18	4.62	p <0.005

Table Ib. Results of the grouped data t-test (one tailed) for the peak intraoral air pressure for [p] for any given tone compared pairwise with each other tone for each of the four speakers.

an exception, however, that in the speech of the Mandarin speaker the mean peak intraoral air pressure in [p] associated with both 35 and 214 is the same, which is 3.56 cm H_2O, although the values of the mean F_0 onset for these two tonal categories are 140 Hz and 99 Hz respectively.

For each speaker, the peak intraoral air pressure for [p] for any given tone was compared pairwise with each other tone using grouped data t-tests (one tailed). The results, shown in Table Ib, indicate that the difference between any such two sets of values is significant at the 0.005 level for the Cantonese and the Shanghai speakers. This is also true for the Mandarin speaker, except for the pair which involves tones 35 and 214 referred to above. In the speech of the Chungking speaker, the difference is significant at the 0.025 level between the tone pairs 45 and 13, 21 and 51, and 51 and 13, however, non-significant between the pairs 45 and 21, 45 and 51, and 21 and 13. In other words, only those pairs of tones whose onsets differ by 3 or more in the traditional 5-point notation show a significant difference.

There seems to be some idiosyncracies in the relationship between the peak intraoral air pressure and the F_0 onset for each speaker. For instance, in the speech of the Mandarin speaker, the difference in the mean peak intraoral air pressure is highly significant at the 0.005 level for [p]'s associated with tones 44 and 51, but non-significant for [p]'s associated with tones 35 and 214. However, the difference in the mean F_0 onset is 32 Hz between tones 44 and 51 and 41 Hz between tones 35 and 214. This seems to suggest that for this particular speaker a greater difference in F_0 onset is not necessarily the result of a significant difference in peak intr oral air pressure. In this case, the level of the F_0 onset seems to have a role in determining when a difference in F_0 onset is associated with a significant difference in the peak intraoral air pressure. The mean F_0 onsets of tones 44 and 51 are at a higher pitch level (248 Hz and 216 Hz respectively) than those of 35 and 214 (140 Hz and 99 Hz, or 138 Hz and 104 Hz if only the first data point is taken to be the F_0 onset). Thus, a smaller difference in F_0 onset between two tones may mean a significant difference in peak intraoral air pressure if the F_0 onset of the tones is at a higher pitch level. On the other hand, for the speech of the Chungking speaker a greater difference in F_0 onset between two tones, irrespective of their pitch level, is the factor which determines a significant difference in peak intraoral air pressure. In the speech of both the Mandarin and the Chungking speakers, the difference between peak intraoral air pressures associated with the F_0 onsets of a lower pitch level (35 and 214 for the Mandarin speaker, and 21 and 13 for the Chungking speaker) is not significant. However, this is not so in the speech of the Cantonese speaker. For this speaker, despite that both 33 and 22 have a lower F_0 onset, the difference in peak intraoral air pressure between [p]'s associated with these tones is significant.

In order to determine the degree of association between the values of the peak intraoral air pressure for [p] and the values of the F_0 onset for

Table II.

SPEAKER	df	CORRELATION COEFFICIENT (r)	SIGNIFICANCE
Mandarin	35	0.800	0.01
Chungking	35	0.536	0.01
Cantonese	28	0.678	0.01
Shanghai	18	0.803	0.01

Table II. Results of the correlation analysis for the values of the peak intraoral air pressure for [p] and the values of the F_0 onset for the following vowel or diphthong.

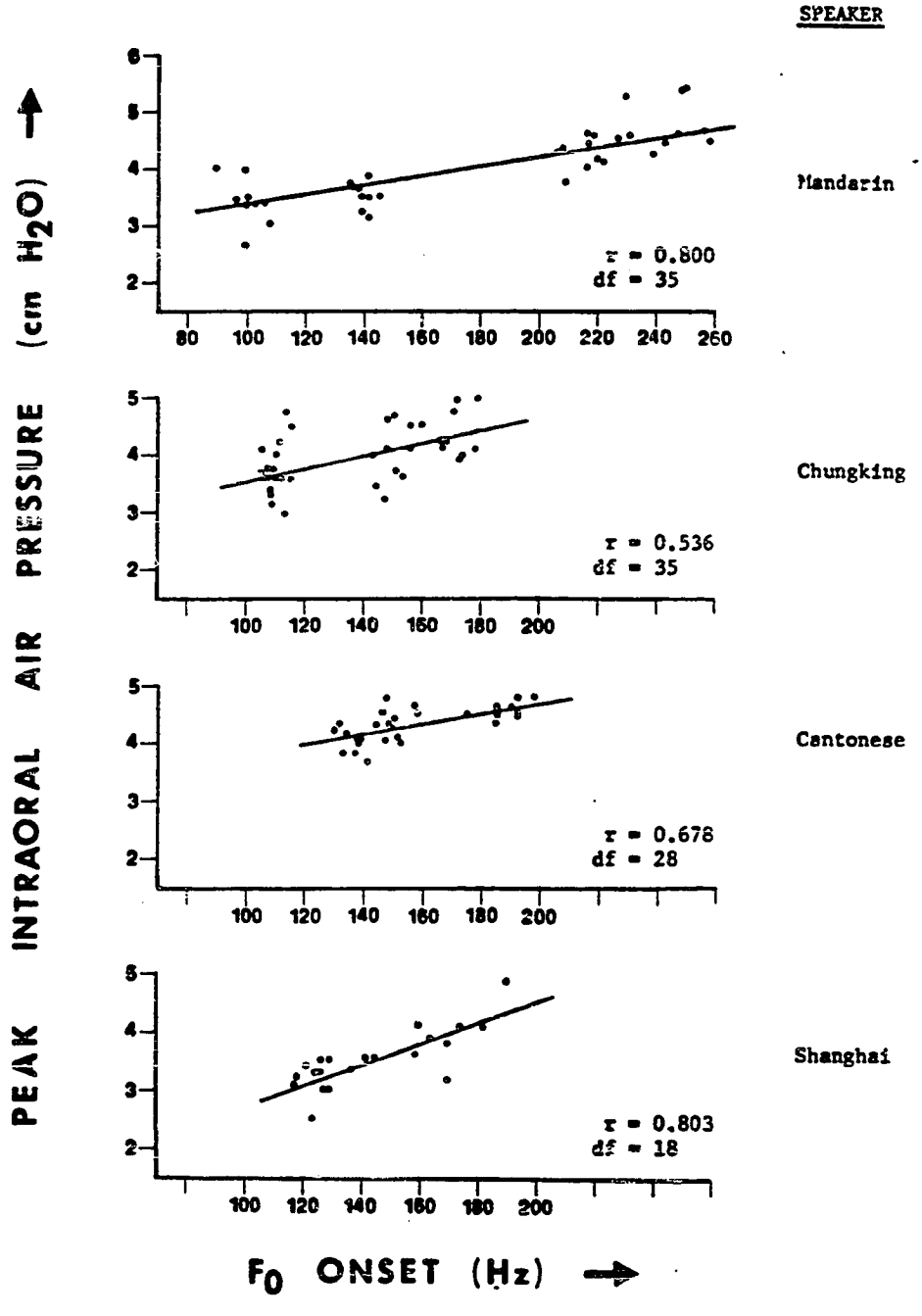


Figure 1. Scatter diagram for the peak intraoral air pressures plotted as a function of F_0 onset variation for each of the four speakers.

the following vowel or diphthong, a Pearson correlation analysis was performed for the speech of each of the four speakers. The results, shown in Table II, indicate that there is an overall positive correlation between the two sets of values (significant at the 0.01 level) for all four speakers. Also shown in the table are the values of correlation coefficient (r). Notice that the correlation coefficient for the Chungking speaker is only 0.536 which is the smallest compared to the r values for the other speakers. This is basically due to the high variance in the Chungking data. As shown in Table Ia, for the Chungking speaker, the standard deviations of the values of the peak intraoral air pressures in [p]'s associated with tones 21 and 45 are 0.60 and 0.51 respectively which are quite large compared to the values of the other standard deviations. Thus, in the speech of the Chungking speaker the peak intraoral air pressures in [p]'s do not correlate well with the F_0 onsets of the two tones, despite a good overall correlation.

Our findings are summarized by four scatter diagrams, as shown in Figure 1, for the speech of each of the four speakers, in which the peak intraoral air pressures are plotted as a function of F_0 onset variation. Also presented in the figure is a linear regression line in each of the four diagrams. Our findings indicate that in the speech of the four speakers the peak intraoral air pressures in [p]'s differ according to the F_0 onset of the tone on the following vowel or diphthong.

4. DISCUSSION

Despite the fact that a higher F_0 is usually accompanied by a higher subglottal pressure (Ladefoged, 1963, Öhman and Lindqvist, 1966), it has been shown that changes in subglottal pressure cannot be the only contributing factor to F_0 variation (Ladefoged, 1963, Öhman and Lindqvist, 1966, Ohala, 1973, 1978). Our findings support the hypothesis that pitch variations are attributed to factors other than subglottal pressure, that is, to laryngeal factors. We, of course, have presupposed that the peak intraoral air pressure and the subglottal pressure are equivalent during the production of a voiceless stop consonant (Ladefoged, 1967, 1968, Netsell, 1969, Scully, 1969). We have shown that in our study the peak intraoral air pressures have a good correlation with the F_0 onsets of the tone on the following vowel or diphthong in most cases. This suggests that the onset subglottal pressure is higher for a tone with a higher F_0 than for a tone with a lower F_0 onset. However, the effect of subglottal pressure on F_0 has been found to be approximately from 2.5 Hz/cm H_2O (Ohala and Ladefoged, 1969, Hixon, Mead and Klatt, 1971) to 2.9 Hz/cm H_2O (Ohala 1978, based on Ohman and Lindqvist, 1966). In our results the mean F_0 onset variations (Table Ib) are far too great to be entirely attributed to the changes in subglottal pressure. The difference between the mean peak intraoral air pressures for any pair of tones never exceeds 1.5 cm H_2O, e.g., in the speech of the Mandarin speaker, the mean F_0 onset differences between tones 44/214, 35/51, and 214/51 are 117 Hz, 108 Hz and 148 Hz respectively, whereas the corresponding differences in the mean peak

intraoral air pressure are merely 0.72 cm H_2O, 1.20 cm H_2O, and 1.20 cm H_2O. Furthermore, in the speech of the same speaker, the difference in the mean F_0 onset between the tones 35 and 214 is 41 Hz, however, there is no difference in the mean peak intraoral air pressure in [p]'s associated with these two tones. These cases indicate that the difference in subglottal pressure cannot account for the large differences in F_0. Similar cases are also found in the speech of the other three speakers (see Table Ib), although the differences in the mean F_0 onset are in a smaller magnitude. Furthermore, as we have described there are high variances in the peak intraoral air pressure in [p]'s associated with tones 21 and 45. This shows that for some tones this speaker does not use pressure to control F_0. Thus, our results provide evidence in support of the hypothesis that variation of F_0 cannot be principally attributed to change in subglottal pressure. Although the changes in subglottal pressure cannot possibly account for the large variations of F_0 onset, the good correlation between the F_0 onsets and the peak intraoral air pressures in [p] nevertheless indicate that the changes in subglottal pressure that accompany the F_0 variations may be a facilitating, but not a necessary, physiological effort for manipulating pitch change in natural speech.

The issue of tone effects on consonants has been controversial. Hyman (1973b) and Hyman and Schuh (1974) denied that such an effect is possible. However, Maddieson (1978) has shown that consonants are affected by tone in terms of diachronic changes of consonants, synchronic difference in consonants and in terms of small but measurable phonetic differences. Our results are in agreement with Maddieson (1978) as far as the phonetic effect of tone on consonant is concerned, as we have shown how the peak intraoral air pressure in a voiceless unaspirated stop is affected by the F_0 onset of the vowel or diphthong that follows the consonant.

5. CONCLUSION

We have shown that (1) there is a good correlation between the peak intraoral air pressure in [p] and the F_0 onset of the tone on the following vowel or diphthong; (2) the amount of change in subglottal pressure cannot account for the large variation of F_0; (3) in terms of peak intraoral air pressure, a consonant is affected by tone.

REFERENCES

Arkebauer, H.J., Hixon, T.J. & Hardy, J.C. (1967). Peak intraoral air pressures during speech. J. of Speech and Hearing Research 10, 196-208.

Bernthal, J.E. (1978). Intraoral air pressure during the production of /p/ and /b/ by children, youths and adults. J. of Speech and Hearing Research 21, 361-371.

Black, J.W. (1950). The pressure component in the production of stop consonants. J. of Speech and Hearing Disorders 15, 207-210.

Brown, W.S. and McGlone, R.E. & Profitt, W.R. (1973). Relationship of lingual and intraoral air pressure during syllable production. J. of Speech and Hearing Research 16, 141-151.

Hixon, T.J., Mead, J. & Klatt, D.H. (1971). Influence of forced transglottal pressure changes on vocal fundamental frequency. JASA, 49, 105.

Hyman, L.M. (1973b). Consonant Types and Tone (Southern California Occasional Papers in Linguistics 1). University of Southern California, Los Angeles.

Hyman, L.M. & Schuh, R. (1974). Universals of tone rules: evidence from West Africa. Linguistic Inquiry 5, 81-115.

Ladefoged, P. (1963). Some physiological parameters of speech. Language and Speech 6, 109-119.

Ladefoged, P. (1967). Three Areas of Experimental Phonetics. London, Oxford University Press.

Ladefoged, P. (1968). Linguistic aspects of respiratory phenomena. Annals of the New York Academy of Sciences 155, 141-151.

Lisker, L. (1970). Subglottal air pressure in the production of English stops. Language and Speech 13, 215-231.

Löfqvist, A. (1971). Some observations on supraglottal air pressure. Phonetics Laboratory, Lund University, Working Papers 5, 27-44.

Maddieson, I. (1978). Tone effects on consonants. J. of Phonetics 6, 327-344.

Netsell, R. (1969). Subglottal and intraoral air pressure during the intervocalic contrast of /t/ and /d/. Phonetica 20, 68-73.

Ohala, J.J. (1973) The physiology of tone. In L. Hyman (Ed.), Consonant Types and Tone. Southern California Occasional Papers in Linguistics 1, 1-14.

Ohala, J.J. (1978). Production of tone. In V. Fromkin (Ed.), Tone: a Linguistics Survey. N.Y., Academic Press, 5-32.

Ohala, J.J. & Ladefoged, P. (1969). Further investigation of pitch regulations in speech. UCLA Working Papers in Phonetics 14, 12-24.

Öhman, S. & Lindqvist, J. (1966). Analysis-by-synthesis of prosodic pitch contours. Speech Transmission Laboratory, Stockholm, QPSR 4/1965, 1-6.

Ringel, R., House, A & Montgomery, A. (1967). Scaling articulatory behavior: intraoral air pressure. JASA 42, 1209A.

Scully, C. (1969). Problems in the interpretation of pressure and air flow data in speech. University of Leeds, Phonetics Dept. Report 2, 53-92.

Subtelny, J.D., Woth, J.H. & Sakuda, M. (1966). Intraoral pressure and rate of flow during speech. J. of Speech and Hearing Research 9, 498-518.

Warren, D.W. (1964). Velopharyngeal orifice size and upper pharyngeal pressure-flow patterns in normal speech. Plast. Reconst. Surg. 33, 148-162.

Section 6: *Tones and Tone Sandhi in Shanghai: Phonetic Evidence and Phonological Analysis*

Tones and tone sandhi in Shanghai: phonetic evidence and phonological analysis

0. ABSTRACT.

It has been claimed that tone sandhi in Shanghai Chinese consists only of the rightward spreading of the tone on the first syllable of a bisyllabic or a polysyllabic compound over the whole compound. To determine if the pitch contours on monosyllables and on compounds are actually similar the pitch patterns of monosyllabic words and of compounds containing two, three or four syllables have been measured for one speaker. The findings show that tone spreading will partially explain the patterns of tone contours, on the compounds, but that there are also some more arbitrary processes involved in tone sandhi in Shanghai. A phonological analysis is presented which regards contours as sequences of level tones, and tones are treated as having an 'autosegmental' association with syllabic units.

1. INTRODUCTION.

Tone sandhi in Shanghai occurs in various types of compounds which are derived by combining monosyllabic words[1]. This kind of compound formation is a highly productive process. For instance, different types of compounds may be derived by combining two or more of the following four monosyllabic words (1) in various different orders (tones are transcribed here according to the system which will be justified in later sections of this paper:

(1)	/tsɔ/	/MH/	'to illuminate'
	/ɕiã/	/MH/	'symbol; object'
	/tɕi/	/HL/	'machine'
	/ɕiɔ/	/MH/	'small'

1 Certain words which are like these compounds in form may not be amenable to a division into constituent morphemes,e.g., certain loan words, and fossilized compounds whose historical origin is forgotten. These may however be treated analogously to those compounds formed by the productive process illustrated here.

Some of the possible compounds which may be found from this small lexicon are given in (2-4).

(2) Bisyllabic compounds:

```
/MH/      /MH/                M  M↑
/tsɔ/  +  /çiã/   ----→  [tsɔ çiã]
'to illuminate'  'symbol'   'photograph'

/MH/      /MH/                M   M↑
/çiɔ/  +  /tsɔ/   ----→  [çiɔ tsɔ]
'small'  'to illuminate'  'portrait
                           (photograph)'
```

(3) Trisyllabic compounds:

```
/MH/      /MH/      /HL/              M   H   L
/tsɔ/  +  /çiã/  +  /tçi/   ----→  [tsɔ çiã tçi]
'to illuminate''symbol' 'machine'      'camera'

/MH/      /MH/      /MH/              M   H   L
/çiɔ/  +  /tsɔ/  +  /çiã/   ----→  [çiɔ tsɔ çiã]
'small' 'to illuminate''symbol'      'small photograph'
```

('↑' = tone raised)

(4) Quadrisyllabic compound:

```
/MH/      /MH/      /MH/      /HL/              M   H   M   L
/çiɔ/  +  /tsɔ/  +  /çiã/  +  /tçi/   ----→  [çiɔ tsɔ çiã tçi]
'small' 'to illuminate''symbol''machine'      'small camera'
```

The most extensive previous discussion of tone sandhi in Shanghai seems to be that of Sherard (1972). Sherard provides some impressionistic phonetic data on the pitch contours observed on some compounds. He proposed the following two generalizations about the patterns he found. (1) The tone contour over an entire bisyllabic or polysyllabic compound is dependent on the tone type of the first syllable, although the resultant contours do not necessarily have the same shape of that of the tone on the first syllable, and (2) bisyllabic and polysyllabic compounds with a first syllable of the same tone type have the same tone contour.

Ballard (1976), basing himself on Sherard, stated that "in Shanghai apparently the only sandhi that occurs is of the type I shall call 'right spreading'. Within a phonological word (or, a compound), all syllables after the first one ignore their citation tone and take their tone from the tone of the first (leftmost) syllable by its spreading out over the whole phonological word." (p. 12) In other words, Ballard suggests that the pitch contours on compounds do necessarily have the same shape as those on their first syllables in isolation. In fact, most of the data presented by Sherard is open to such an interpretation but there is relatively little discussion of compounds of more than two syllables.

What is of especial interest in these accounts is the suggestion that the contours are treated as unitary wholes. If the monosyllabic contour is extended as a unit over the longer compound this would represent an important piece of evidence in favor of the proposition that tone contours are in fact units in some languages (see Anderson, 1978, for some discussion of the issue). Therefore the reality of such a rule in Shanghai would have an important consequence for phonological theory. Kennedy (1953) proposed a similar rule for his native Wu dialect of Tangsi but the question of the unity of contours was not explicitly raised and it cannot be resolved from the data he cites. More extensive documentation from relevant languages is needed: this paper is offered as a contribution to that process with respect to Shanghai.

The study reported here is concerned first with comparing the shapes of pitch contours on various classes of monosyllabic words with those on an extensive variety of compound words beginning with monosyllables of different classes. Exact pitch contours have been obtained so that this comparison may be based on objective criteria. Because only one speaker was used for these comparisons, appropriate caution must be used in drawing generalisations of broader scope. However, the acoustic analysis offers a solid foundation for constructing a detailed phonological account of the processes involved in deriving the tone patterns on compounds in the speech of this subject. In particular it makes it possible to determine how far a 'tone-spreading' explanation of the sandhi processes in compounds can account for the facts of the case, and helps to clarify the issue of whther the contours on monosyllables are being spread out as indivisible units.

2. TONES ON THE MONOSYLLABIC WORDS.

Five of the eight etymologically distinct 'tones' of Middle Chinese (c.600 A.D.) remain distinct in contemporary Shanghai. Although these are traditionally referred to as tones, the differences between them concern several properties of the syllable, including the pitch, the presence or absence of a final glottal stop, the presence or absence of voicing in initial

obstruents and the duration of the syllable. It is therefore more accurate to talk of five syllable types. Three of these are long syllables with contrasting pitch patterns; these syllables are open or nasal final. The remaining two syllable types are short and have a final glottal stop but contrast in pitch. These syllable types will be identified by the letters, A, B, C, D and E. The monosyllabic words in Table I show the contrast between these types. Note that initial obstruents in syllables

	SYLLABLE TYPE	MONOSYLLABIC WORD	
Long, Open	A	[pî]	'edge'
	B	[pī]	'change'
	C	[bǐ]	'skin'
Short, Closed	D	[pɪ́ʔ]	'pen'
	E	[bɪ̀ʔ]	'other'

initial voicing with obstruents

[´ = HIGH; ¯ = MID; ` = LOW; ˆ = ´`; ˇ = `´; ˩ = ` –]

TABLE I. Syllable types and monosyllabic words.

of types C and E must be voiced, obstruents elsewhere are voiceless. The words in Table I were used for the acoustical analysis of pitch on monosyllabic words.

A word list containing 50 tokens of these words (5 x 10 repetitions), arranged in a random order, was prepared. Each word was placed in the carrier frame as shown below:

[ŋū	dɔ̀ʔ	____	pʌ́ʔ	nóŋ	tʰîŋ]
I	read	____	give	you	listen
'I	read	____	for you to listen.'		

The speaker used in this investigation was female native of Shanghai. She is in her late fifties and was once a high school teacher in Shanghai. Recently she immigrated into the United States via Hong Kong where she lived no more than five years. Her speech was judged by other native

Shanghai speakers, including the first author, and was considered to be standard metropolitan Shanghai. The speaker was instructed to read the word list at a normal rate of speech. The recording was performed in a sound treated room. The tape was analyzed using the PDP-12 computer at the UCLA Phonetics Lab. Fundamental frequency measurements were derived every 10 msec by the CEPSTRUM method . Duration measurements of the vowels were also made on the waveforms displayed on the computer screen. Because the long syllables differ substantially in duration the data from the pitch analysis for these syllables was normalized in the time domain in the following manner. The measurements of the fundamental frequency for each token were divided into five sections, each containing as nearly as possible an equal number of 10 msec intervals and the means of the values of the 10 msec intervals within each section were calculated. A mean value for each of the five sections of each token was thus obtained. These values were then averaged across tokens of the same word to obtain means for each word. These mean values are represented by the points shown in Figure 1. No such normalization process has been applied to the pitch data for the short syllables, as the duration of these syllable types is so short that normalization would severely distort the true picture of the pitch contour.

The phonetic pitch shapes of the five syllable types are shown in Figures 1 and 2. Each contour on Figure 2 represents one single token. Duration is not normalized on Figure 2. Figure 1, however, shows the pitch of syllable types A, B and C of normalized duration, and each is the average of 10 tokens. Figure 3 shows the pitch contours of three additional tokens of type D and E syllables. From these figures, we see that a type A syllable has a high falling pitch contour. The pitch contour of a type B syllable is fairly level, but has a slightly rising movement for most of its duration. A syllable of type C has a rising contour. The pitch contour of a type D syllable starts almost as high as type A and remains at this level for a short while and then falls sharply. As for type E syllables, they rise, but then fall sharply at the end.[2]

So that comparisons of length may be made the duration (in msec) of the vowels in the five syllable types is given in Figure 4. Each value on top of the dark bars is the average of 10 tokens.

These findings generally agree with previous descriptions of the pitch contrasts that occur on the monosyllabic words. There are, however, some minor differences in the phonetic descriptions of the pitch shapes. JIANGSUSHEN HE SHANGHAISHI FANGYAN GAIKUANG (JHSFG: Synopsis of Shanghai Dialect and the Dialects in Jiangsu Province, 1960) describe the pitch contours of the five syllable types according to a 5-point numerical scale, where 5 = high, as follows:

2 Sokolov (1975) presented some measurements of the initial and final fundamental frequency contours of monosyllabic words spoken by one male and one female speaker of Shanghai. His data appear not to be inconsistent with the measurements presented here.

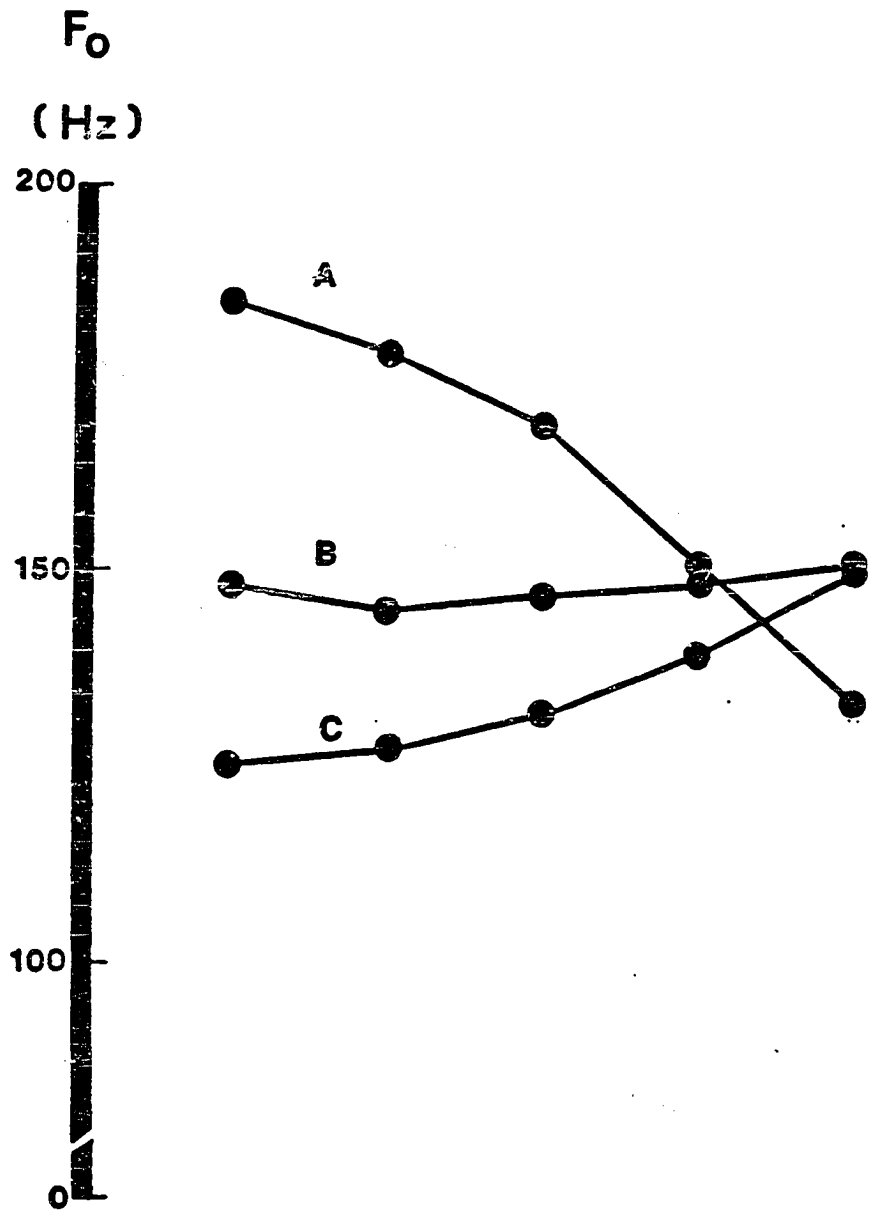


FIGURE 1. Pitch contours of normalized duration of syllable types A, B, and C.

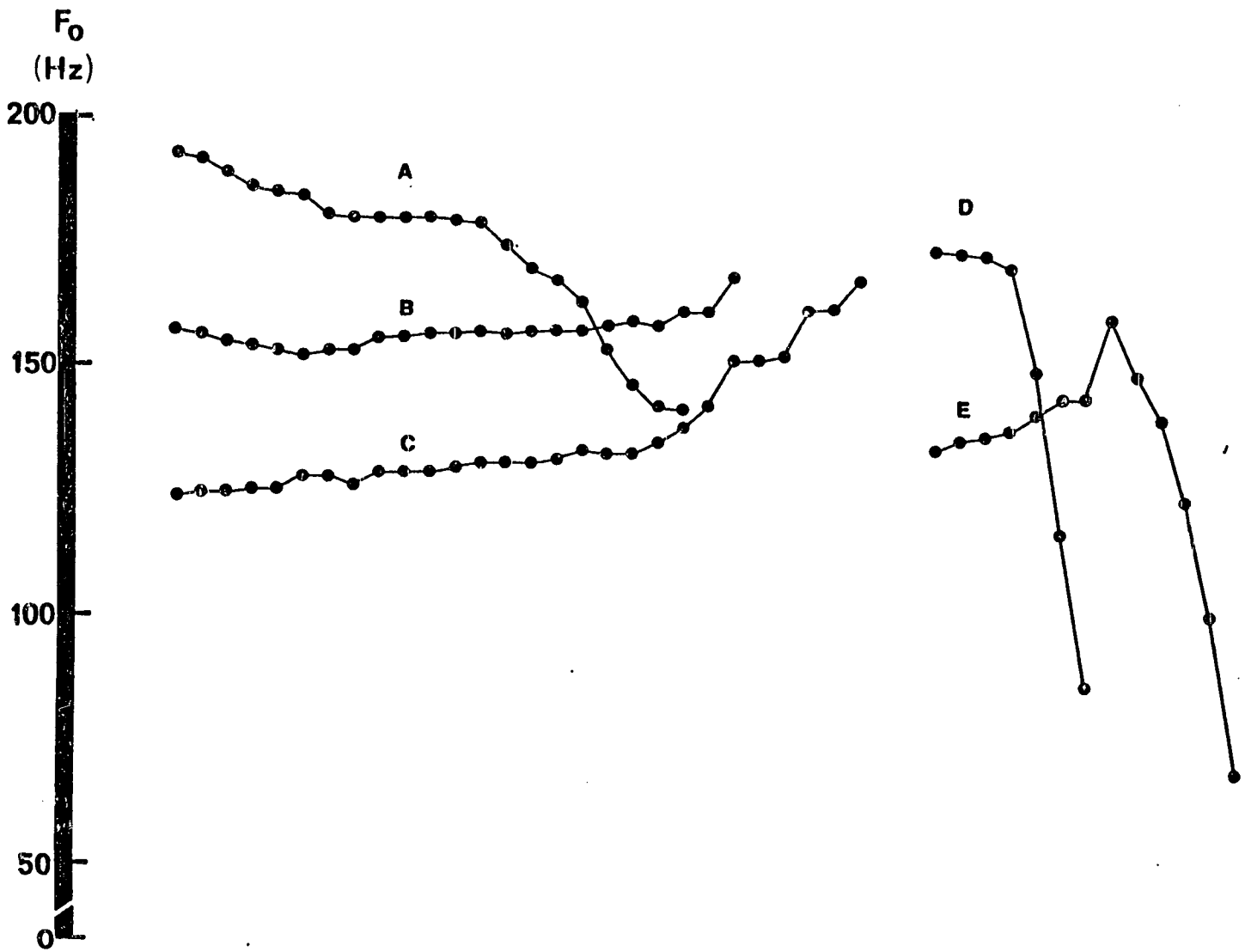


FIGURE 2. Pitch contours on syllable types A, B, C, D and E.

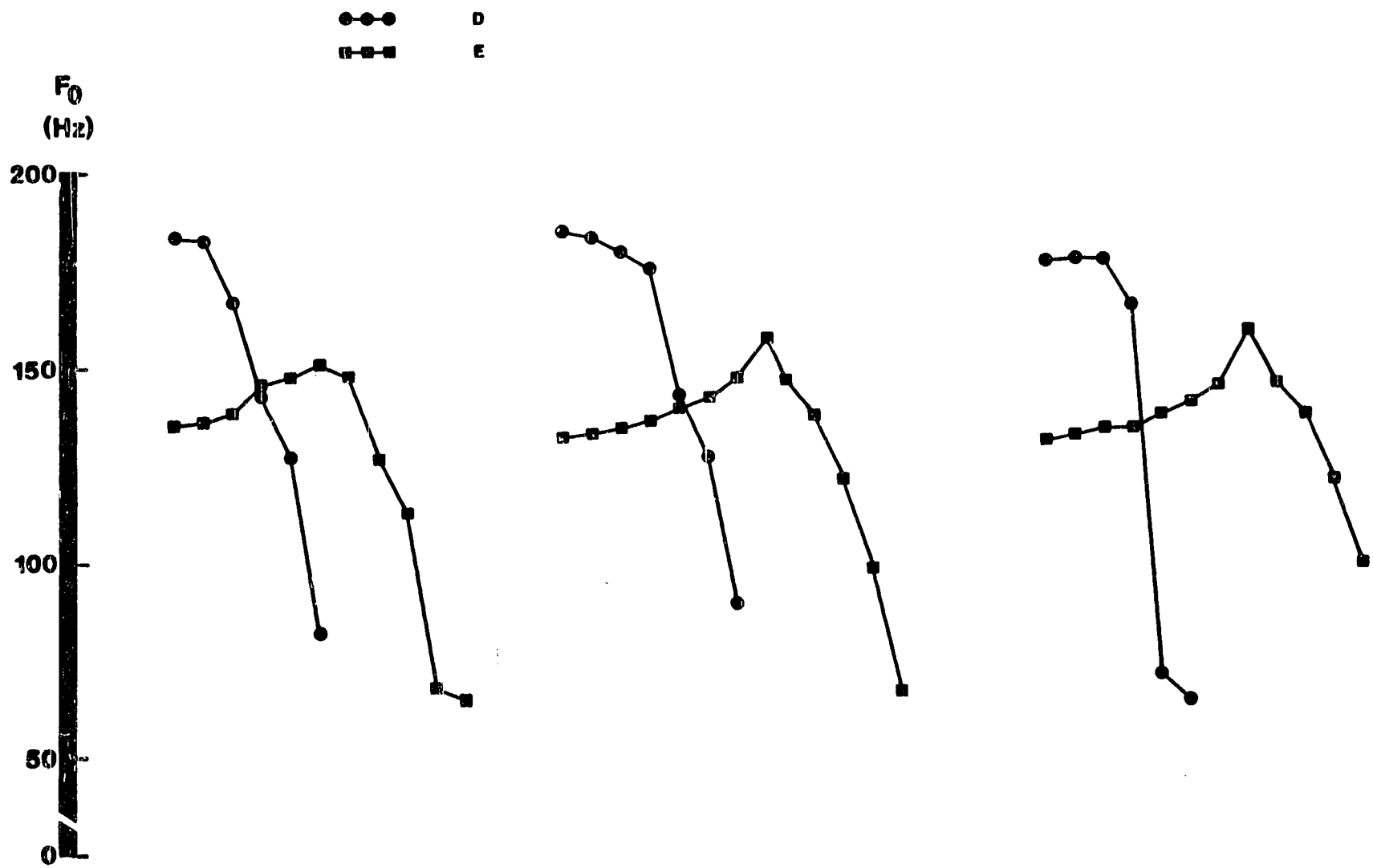


FIGURE 3. Pitch contours of syllable types D and E.

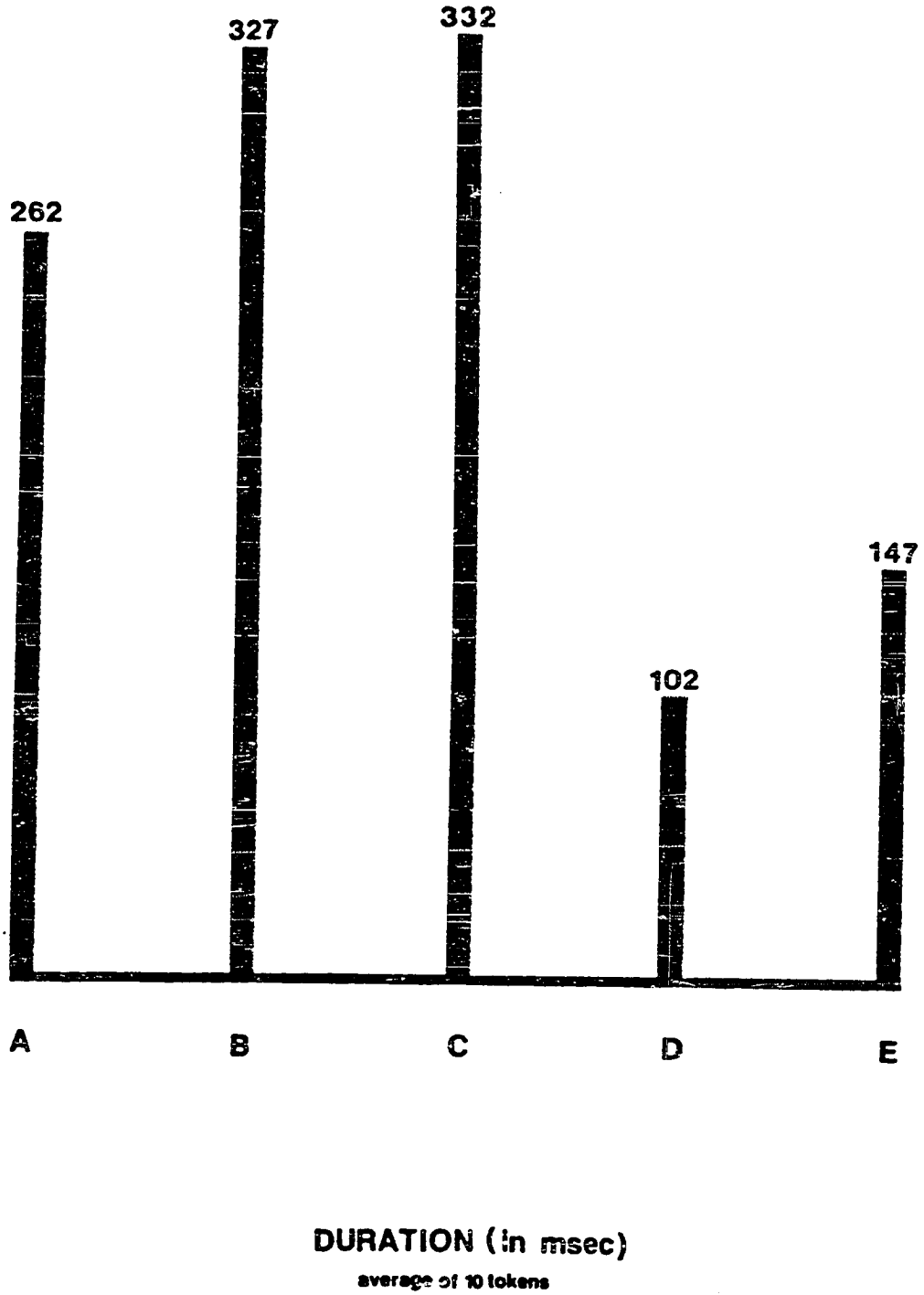


FIGURE 4. Durations (in msec) of syllable types A, B, C, D and E.

	SYLLABLE TYPE	JHSFG	HFG
Long	A	53	42
	B	34	35, 434
	C	14	24
Short	D	5	5
	E	2	23

Sherard (1972) offers a qualitative description of the pitch shapes, in which syllables of type A are characterized as having a sharp falling tone, B as having a moderately high level tone with no discernible peak in pitch height in normal speech, C as having a tone in low register with clearly audible tone contour with end point higher than the onset, D as having a tone of moderately high pitch with no discernible change of pitch, and E as having a tone in low register throughout with a short rise in pitch. The differences among the impressionistic studies, as far as the pitch shape of monosyllabic words is concerned, are in the description of syllables of types B and E.

The pitch of type B syllables is described by JHSFG as mid with a slight rise (34 in numerical terms). HFG implies that it is more sharply rising (35), or is in a higher-mid range but has a dip in the middle (434). Sherard describes it as moderately high and basically level but he draws it with a slightly dipping contour. Our acoustical data show that the pitch at the onset of type B syllables is midway between the pitch at the onset of types A and C. The pitch dips slightly after the onset and then rises for most of its duration. The end point is higher in pitch than the onset. We are inclined to believe the characterization of this contour as basically mid-rising is correct. The tendency to fall after the onset which produces the dipping pattern probably can be related to the fact that basically level tones actually fall in most languages, especially on longer syllables. Apparently JHSFG did not observe this initial fall, whereas both HFG and Sherard paid some attention to it. The amount and to some extent the direction of pitch movement observed seems to vary but some rising movement is always reported.

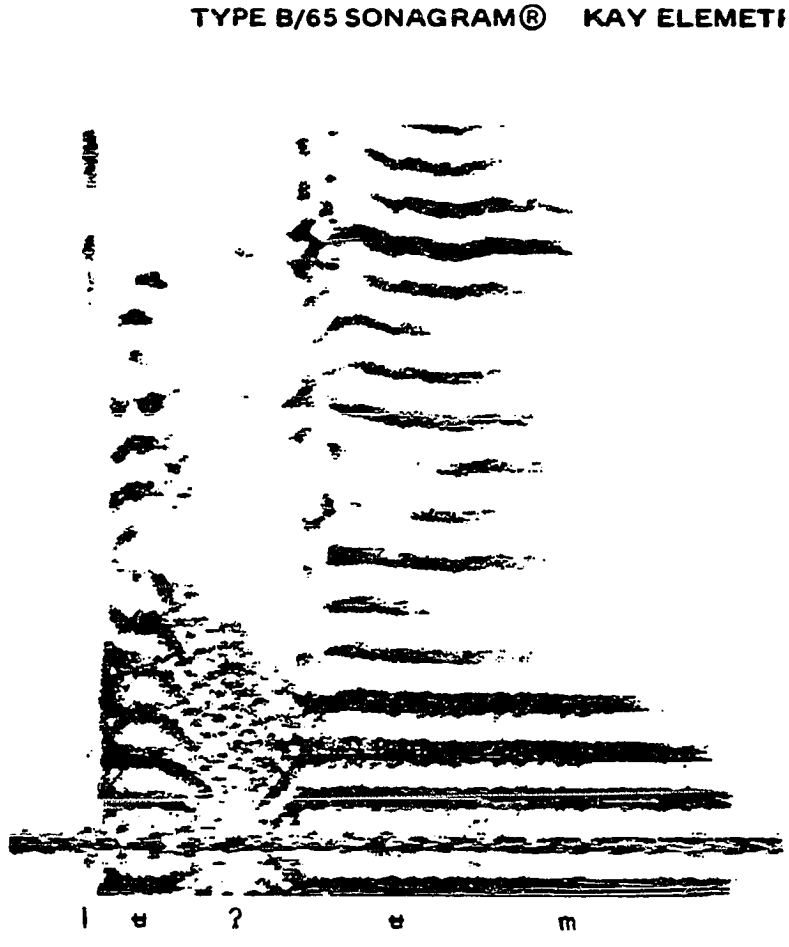
The pitch on type E syllables has been described as either short low rising (HFG and Sherard) or as short low level tone(JHSFG). Our acoustical data (Figure 2) show that the final portion of the pitch contours of both type D and type E syllables falls to an extremely low

point for our subject. This sharp fall has not been reported in the previous studies, but we are doubtful if this means it was missing from the speech of their subjects. The authors may not have perceived its presence due to its extreme shortness (in our subject's speech this falling pitch movement is extremely difficult to perceive by ear). If it is noticed at all, it may rather be interpreted as a cue for the presence of the final stop. Recall that both type D and type E syllables end with a syllable final glottal stop, (C)Vʔ. We believe that the sharp falling portion of the pitch contours is caused by the closing action of the vocal cords for the glottal stop. If this is true we would expect to find similar effects in other languages, and indeed, a similar case is found in !Xõ, a language spoken in South West Africa (Traill, in progress). In !Xõ, the final portion of the pitch contour of a vowel drops sharply when the vowel is followed by a glottal stop. Figure 5 shows a narrow band spectrogram of the pitch contour of the word [ǀʉʔʉm] "to cool down". We can see that the portion of the pitch contour that precedes the intervocalic glottal stop falls sharply.

Examining the pitch contours of type D and type E syllables in Figure 1, we see that the pitch begins to drop sharply at the 5th time point for type D syllables and at the 9th time point for type D syllables and at the 9th time point for type E syllables. We can reasonably assume that closure for the glottal stop begins to be articulated at these time points. Since we argue that the sharp fall in the pitch contours of type D and type E syllables is due to segmental influence, that is, it is a property that can be predicted in this environment, we exclude this property from the phonetic representation of the pitch contours of the short syllables. Thus,phonetically we may describe these pitch contours as high level (type D syllables) and low rising (type E syllables) for our subject. This characterisation of type D agrees with all the other sources, and this description of type E agrees with HFG and Sherard. The speech variety reported in JHSFG may have differed in having a level pitch pattern in type E syllable, but perhaps this source has merely used a simplified transcription for the two-way contrast in short syllables.

Based on our acoustical data and the above discussion,we summarize the essential pitch characteristics of monosyllables of the types A, B, C, D and E in Shanghai as follows:

(1) type A syllables - start high then fall to low.
(2) type B syllables - start at a mid level and rise slightly.
(3) type C syllables - start low then rise to the same level as the end of type B.
(4) type D syllables - high level.
(5) type E syllables - start low then rise to almost the same level as the end of type B and type C.



[lʉʔʉm] 'to cool down'

(Language: !Xõ:)

FIGURE 5. Narrow band spectrogram of [lʉʔʉm].

Note that the pitch contours of type C and type E syllables are similar, but that type B starts at a higher level than C or E. This difference in the contours is much too great to be attributed to the possible phonetic effect of the initial consonants in the test items used with our subjects. The pitch levels in B and C type syllables remain sharply different even as long as 250 msec after the vowel onset, so that even though the type B word measured begins with [p] and the type C word measured begins with [b] the voicing difference is not plausibly responsible for the different contours on these syllables. (See Hombert, 1978, for more discussion on the extent of consonant perturbations of F_o).

3. PITCH CONTOURS OF THE BISYLLABIC, THE TRISYLLABIC AND THE QUADRISYLLABIC COMPOUNDS.

For comparison with the contours on monosyllables pitch contours for bisyllabic, trisyllabic and quadrisyllabic compounds were obtained using the same subject and measurement technique described above. The compounds chosen for analysis and the members within the compounds are listed in Tables II-IV. The compounds are composed of combinations of elements from the five etymologically distinct tone types in all possible combinations of two syllable types and selected sequences of three and four different types.

The possible pitch contours in bisyllabic compounds, and the pitch contours of the selected trisyllabic and the quadrisyllabic compounds are shown in Figure 6, Figure 7 and Figure 8 respectively. They are actual copies of the fundamental frequency contours displayed on the computer screen. The syllable type of each of the syllables in these compounds is indicated by A, B, C, D or E. Each dot, large or small, represents the fundamental frequency value at a certain time point, (10 msec apart). The large dots also mark the beginning and end of the pitch contour of any compound. Each contour represents one single token taken as representative of the type.

Figure 6 contains the pitch contours of the bisyllabic compounds listed in Table II. The data are arranged according to the tonal categories of the isolated monosyllables of which they are constructed. Thus, all compounds in the first row have a monosyllabic word of type A in first position, the second row have a type B syllable in first position, and so on. All the compounds in the first column have a monosyllabic word of type A in second position, the second column have a monosyllabic word of type B, and so on. As we can see, the shapes of the pitch contours of the bisyllabic compounds within each row (i. e., having the same isolation tone on the first syllable) are broadly similar, but the pitch contours are different within each column. In other words, the pitch contour of the first syllable in isolation largely determines the contour over the compound. Thus, there are five basic pitch patterns for the bisyllabic compounds. However, where the second syllable is a type D syllable and the first is

BISYLLABIC COMPOUNDS

#1 /tʰî/ 'day' + /tʰî/ 'day' ➡ [tʰítʰì] 'everyday'

#2 /tʰî/ 'sky' + /tʰɕǐ/ 'air' ➡ [tʰítʰɕì] 'weather'

#3 /tʰî/ 'sky' + /dǐ/ 'earth' ➡ [tʰídì] 'universe'

#4 /pî/ 'to arrange' + /tʰɕí?/ 'to compile' ➡ [pítʰɕí?] 'editor'

#5 /fî/ 'flying' + /dǐ?/ 'saucer' ➡ [fídí?] 'flying saucer'

#6 /tɕǐ/ 'to cut' + /tɔ̂/ 'knife' ➡ [tɕītɔ̄] 'sissors'

#7 /tɕǐ/ 'to examine' + /tʰɔ̌/ 'to discuss' ➡ [tɕītʰɔ̄] 'to criticize'

#8 /pǐ/ 'flat' + /dʉ̌/ 'bean' ➡ [pīdʉ̄] 'flat bean'

#9 /tɕǐ/ 'space' + /tɕí?/ 'connection' ➡ [tɕītɕí?] 'indirect'

#10 /tɕǐ/ 'space between' + /dí?/ 'to spy' ➡ [tɕīdí?] 'spy'

#11 /dǐ/ 'electricity' + /tʰî/ 'stairs' ➡ [dītʰī] 'elevator'

#12 /dǐ/ 'electricity' + /tʰɕǐ/ 'tool' ➡ [dītʰɕī] 'electric tool'

#13 /dǐ/ 'younger brother' + /dǐ/ 'younger brother' ➡ [dīdī] 'younger brother'

#14 /dǐ/ 'electricity' + /pí?/ 'pen' ➡ [dīpí?] 'electric heating tube for fish tank'

#15 /dǐ/ 'electricity' + /dʒí?/ 'pole' ➡ [dīdʒí?] 'poles (battery)'

#16 /tɕí?/ 'urgent' + /ɕǐ/ 'need' ➡ [tɕíɕǐ] 'urgent need'

#17 /tɕí?/ 'to connect' + /tɕǐ/ 'to aid' ➡ [tɕítɕǐ] 'to aid'

#18 /tɕí?/ 'knot' + /dʉ̌/ 'head' ➡ [tɕídʉ̌] 'knot'

#19 /tɕí?/ 'urgent' + /tʰɕí?/ 'to cut' ➡ [tɕítʰɕí?] 'urgent'

#20 /tɕí?/ 'to amass' + /dʒǐ?/ 'extreme' ➡ [tɕídʒǐ?] 'motivated'

#21 /zʌ̌?/ 'ten' + /sɛ̂/ 'three' ➡ [zʌ̀sɛ̄] 'thirteen'

#22 /zʌ̌?/ 'day' + /tɕǐ/ 'record' ➡ [zʌ̀tɕī] 'diary'

#23 /zʌ̌?/ 'stone' + /dʉ̌/ 'head' ➡ [zʌ̀dʉ̄] 'stone'

#24 /zʌ̌?/ 'straight' + /tɕí?/ 'to connect' ➡ [zʌ̀tɕí?] 'direct'

#25 /zʌ̌?/ 'sun' + /zʌ̌?/ 'to eat slowly' ➡ [zʌ̀zʌ̄?] 'solar eclipse'

TABLE II. List of bisyllabic compounds and first and second memeber of the compounds.

(´ = H; ¯ = M; ` = L; ˆ = ˋˊ; ˇ = ˊˋ; ˩˥ = ˧˥; '↑' = tone raised)

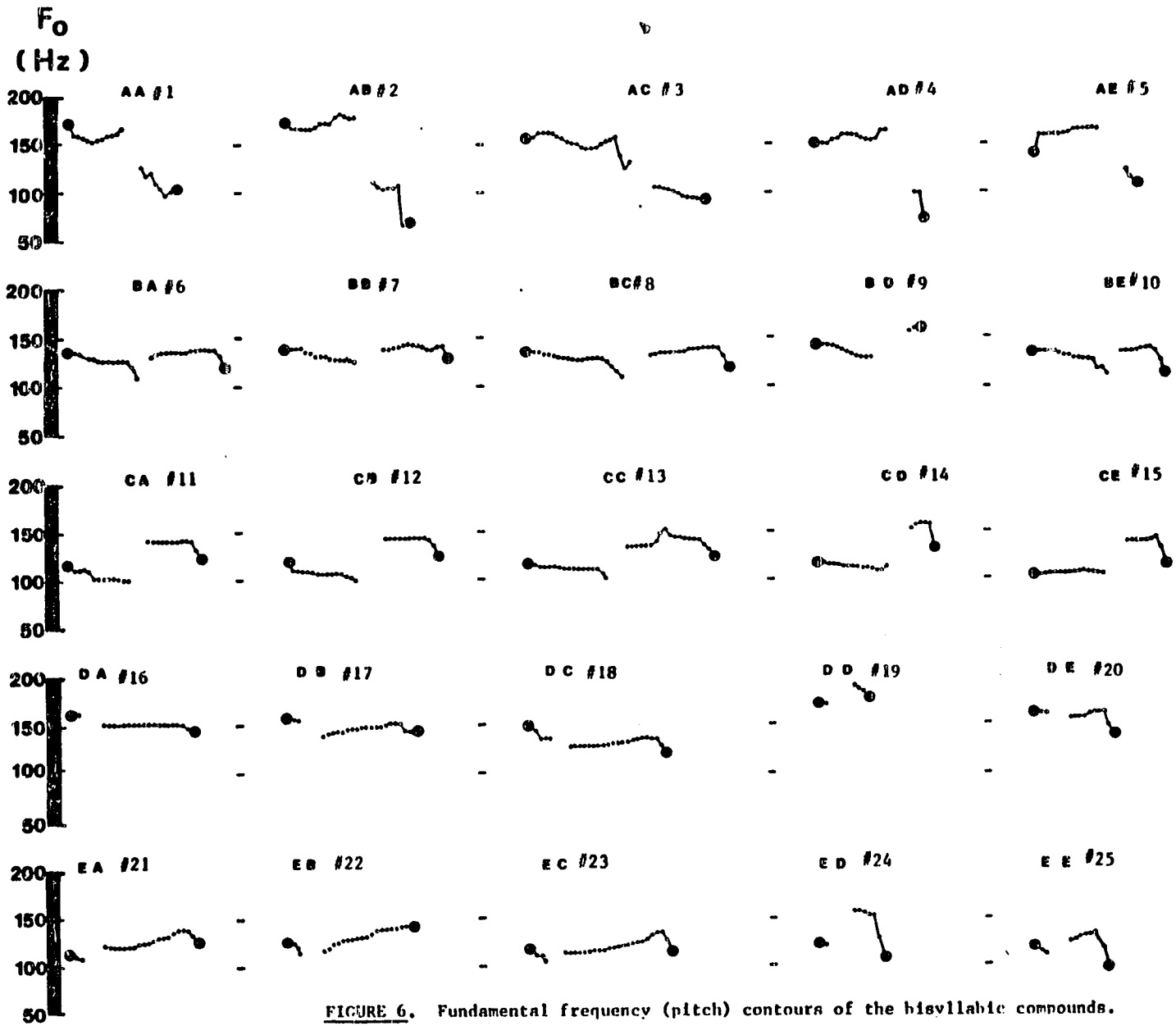


FIGURE 6. Fundamental frequency (pitch) contours of the bisyllabic compounds.

any syllable type except type A, the second portion of the pitch contour is higher than the comparable second portion of the other pitch contours in the same row. With the exception of these cases, the shapes of the pitch contours of the bisyllabic compounds are similar to the shape of the pitch contours of the monosyllabic words of the same type as their first syllable (Cf. Figures 1-3 and Figure 6). Of course, this statement is made based on the presumption that the sharp falling portion of the pitch contours of type D and type E syllables is not an intrinsic tonal property.

Figure 7 contains the pitch contours of the trisyllabic compounds listed in Table III. The data are again arranged in five rows so that all compounds in the first row have a type A syllable in first position, and in the second row have a type B syllable in the first position, and so on. The columns represent compounds with the same syllable type in second position. The third syllable may be any of the five syllable types. We can see that the pitch contours within each row are similar. Thus, there are five patterns of pitch contour that occur on the trisyllabic compounds. Notice that the shape of the pitch contour of the trisyllabic compound with type A or type E syllable as its first syllable is similar to the shape of the pitch contour of type A or type E syllable in isolation. This is however not the case for compounds with type B, C or D syllable as their first syllable. For these trisyllabic compounds, the first portion of their pitch contour is similar to the first portion of the pitch contour of their first syllable when it occurs as a monosyllabic word. The shape of the remaining portion of the contours is similar in all these cases, i.e., it falls to low.

Figure 8 contains the pitch contours of the quadrisyllabic compounds listed in Table IV. One example each is given of a quadrisyllabic compound with type A, B, C, D or E syllable as its first syllable. We can see that the shape of the pitch contour of the compound with type A syllable as its first syllable is similar to the pitch contour of a type A monosyllabic word. However, this is not so for the others. For these cases, the first portion of their contours is similar to the first portion of the pitch contour of their first syllable in isolation. The shape of the remaining portion of the pitch contours is similar in all cases. Note that the pitch pattern of first portion of the quadrisyllabic compounds with type C syllable as their first syllable is similar to the first portion of the compounds with a type E syllable as their first syllable. There are in fact only four pitch contours which occur on the quadrisyllabic compounds.

As may be seen from comparing Figure 7 and Figure 8 the shapes of the pitch contours of the trisyllabic and the quadrisyllabic compounds are generally similar. For example, the pitch contours of the trisyllabic compounds which begin with a type A syllable (the first row in Figure 7) are similar to the contour of the quadrisyllabic compound with an initial type A syllable. The exception to this observation is the compounds whose first syllable is type E. In this case the quadrisyllabic compound has a contour unlike that found on the trisyllabic compounds.

TRISYLLABIC COMPOUNDS

#1 /fɪ̂/ 'flying' + /tɕî/ 'machine' + /sɨ̂/ 'instructor' ➡ [fɪ́tɕîsɨ̀] 'pilot'

#2 /tɕîɔ/ 'to exchange' + /ɕîɪ̃ɔ/ 'sound' + /tʰɕɪɔ́ʔ/ 'song' ➡ [tɕîɔɕîɪ̃ɔ̀tʰɕɪɔ̀ʔ] 'symphony'

#3 /tʰî/ 'sky' + /vɛ̌ŋ/ 'studies' + /dɛ̌/ 'terrace' ➡ [tʰîvɛ̄ŋdɛ̀] 'observatory'

#4 /sɛ̂/ 'three' + /kɔ́ʔ/ 'corner' + /pɛ̌/ 'board' ➡ [sɛ́kɔ̄pɛ̀] 'triangles'

#5 /jîŋ/ 'sound' + /jɪɑ̌ʔ/ 'music' + /tɕîɑ/ 'specialist' ➡ [jîŋjɪ̄ɑ̀tɕîà] 'musician'

#6 /tʰɕî/ 'gas' + /tʰsɔ̂/ 'cart' + /fû/ 'person' ➡ [tʰɕîtʰsɔ́fù] 'chauffeur'

#7 /tsɔ̌/ 'to illumine' + /ɕîɪ̃/ 'symbol' + /tɕî/ 'machine' ➡ [tsɔ̄ɕîɪ̃ɪ̄tɕì] 'camera'

#8 /jî̌ɑ/ 'young' + /zî/ 'juvenile' + /yɛ̌/ 'garden' ➡ [jî̌ɪzîyɛ̀] 'kindergarten'

#9 /sɨ̌/ 'hand' + /tɕî/ 'finger' + /kʰʌ́ʔ/ 'nail' ➡ [sɨ̌tɕîkʰʌ̀ʔ] 'finger nail'

#10 /sɨ̌// 'water' + /mî/ 'honey' + /dɔ̌/ 'peach' ➡ [sɨ̄mîdɔ̀] 'a kind of peach'

#11 /zɔ̌/ 'spirit' + /tɕî/ 'essence' + /bî// 'sickness' ➡ [zɔ̀tɕîbî] 'crazy'

#12 /mɔ̌/ 'horse' + /ɕî/ 'drama' + /dɨ̌/ 'group' ➡ [mɔ̀ɕîdɨ̀] 'circus'

#13 /lɔ̌/ 'old' + /zî/ 'former' + /pɔ̌/ 'generation' ➡ [lɔ̀zîpɔ̀] 'sage'

#14 /xɔ̌ŋ/ 'red' + /ɕîo/ 'blood' + /dʒî̌ɪ/ 'ball' ➡ [xɔ̀ŋɕîodʒî̌ɪ] 'red blood cell'

#15 /dû/ 'big' + /zʌ̌ʔ/ 'gate' + /hɑ̌/ 'crab' ➡ [dûzʌ́hà] 'crabs from Big Gate'

#16 /ɕîʔ/ 'snow' + /hwɔ̂/ 'flower' + /kɔ̂/ 'cream' ➡ [ɕîhwɔ́kɔ̀] 'vanishing cream'

#17 /tʰsʌ́ʔ/ 'to exit' + /kʰɨ̌/ 'mouth' + /tsɛ̌ŋ/ 'document' ➡ [tʰsʌ́kʰɨ̀tsɛ̀ŋ] 'document for leaving a country'

#18 /hʌ́ʔ/ 'black' + /dɛ̌/ 'head' + /fʌ́ʔ/ 'hair' ➡ [hʌ́dɛ̀fʌ̀ʔ] 'black hair'

#19 /pʰɔ́ʔ/ 'to strike' + /kʰʌ́ʔ/ 'competent' + /bɑ̌/ 'placard' ➡ [pʰɔ́kʰʌ́bɑ̀] 'cards (poker)'

#20 /pʌ́ʔ/ 'hundred' + /jî̌ʔ/ 'page' + /tɕîʔ/ 'knot' ➡ [pʌ́jîtɕîʔ] 'a kind of food made of soybeans'

#21 /lɛ̌ʔ/ 'to record' + /jîŋ/ 'sound' + /tɕî/ 'machine' ➡ [lɛ̀jîŋtɕɛ̄î] 'tape recorder'

#22 /zʌ̌ʔ/ 'sun' + /pɛ̌ŋ/ 'original' + /pî̌ŋ/ 'people' ➡ [zʌ̀pɛ̀ŋpî̌ŋ] 'Japanese'

#23 /lɛ̌ʔ/ 'green' + /dɛ̌/ 'bean' + /tʰɔ̌ŋ/ 'soup' ➡ [lɛ̀dɛ̀tʰɔ̌ŋ] 'green bean soup'

#24 /mɔ̌ʔ/ 'eye' + /tîʔ/ '(particle)' + /vʌ̌ʔ/ 'object' [mɔ̀tîvʌ̀ʔ] 'target'

#25 /tʌ̌ʔ/ 'white' + /mɔ̌ʔ/ 'wood' + /pî/ 'ear' ➡ [bʌ̀mɔ̀pî] 'white fungus'

TABLE III. List of trisyllabic compounds and first, second and third members of the compounds.

QUADRISYLLABIC COMPOUNDS

#1 /ɕîŋ/ 'new' + /vɛ̌ŋ/ 'to hear' + /tɕî/ 'to record' + /tsɛ̂/ 'person' ➡ [ɕîŋvɛ̄ŋtɕîtsɛ̀] 'news reporter'

#2 /pɨ̌/. 'half' + /jî̌ɑ/ 'night' + /sɛ̂/ 'three' + /kʌ̌/ 'time measurement' ➡ [pɨ̄jîɑsɛ̄kʌ̀] 'pre-dawn'

#3 /jî̌ɪ/ 'oil' + /tʰɕîɪ̃/ 'behavior' + /wʌ̌ʔ/ 'slippery' + /dî̌ɔ/ 'tune' ➡ [jî̌ɪtʰɕîɪ̃wʌ̄dî̌ɔ] 'frivolous'

#4 /tɕîʔ/ 'to unite' + /hwɔ̌ŋ/ 'matrimony' + /tsɛ̌ŋ/ 'proof' + /sî/ 'book' ➡ [tɕîhwɔ̀ŋtsɛ̄ŋsî] 'marriage licence'

#5 /zʌ̌ʔ/ 'factual' + /zɨ̌/ 'to be' + /dʒî̌ɪ/ 'then' + /zɨ̌/ 'to be' ➡ [zʌ̀zɨ̌dʒî̌ɪzɨ̀] 'unpretentious'

TABLE IV. List of quadrisyllabic compounds and the members of the compounds.

(´ = HIGH; ¯ = MID; ˋ = LOW; and ˆ = ˇ; ˋ = ˋˋ; ˇ = ˋˋ; ˇ = ˋˋ) ('↑' = tone raised; '↓' = tone lowered)

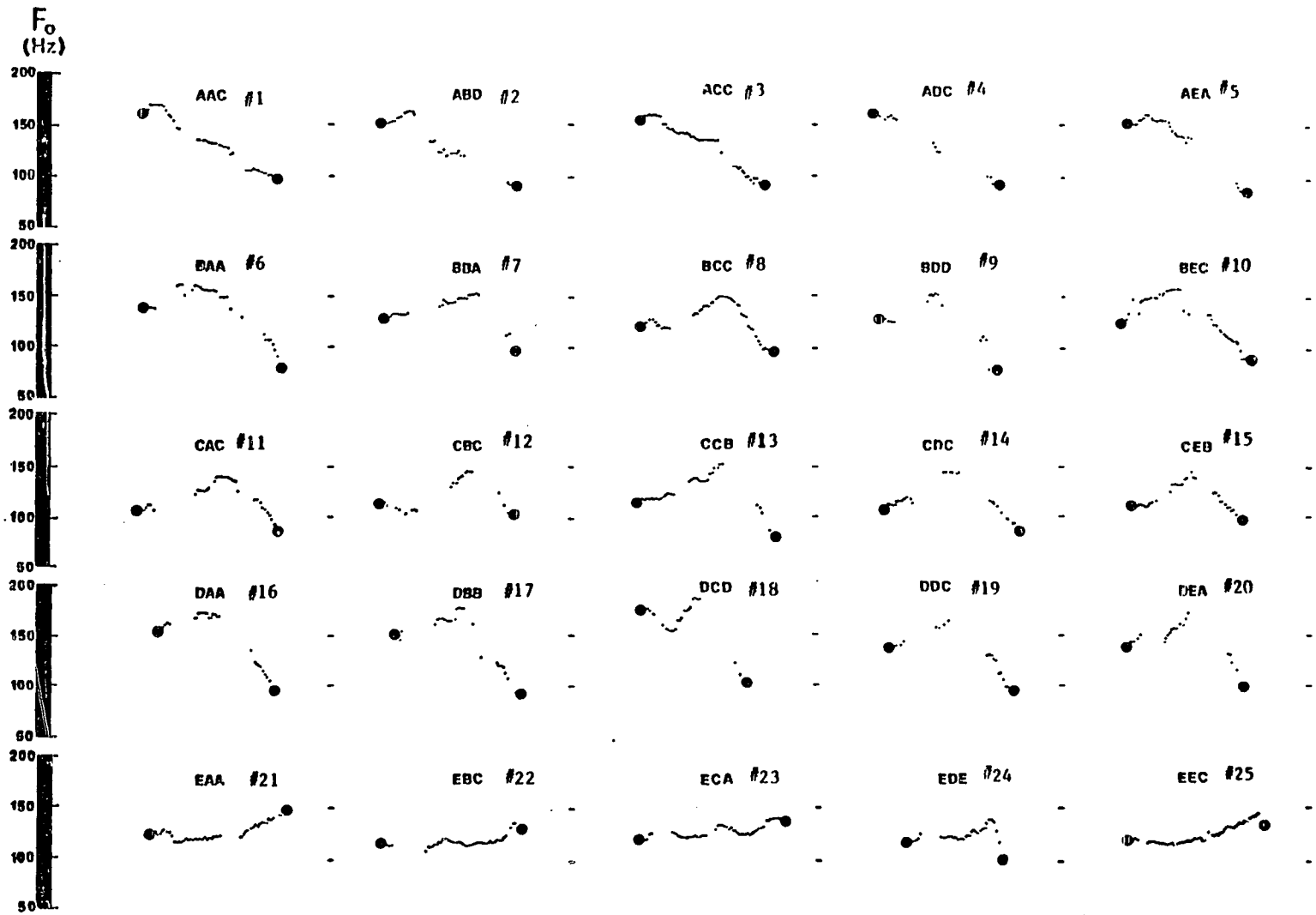


FIGURE 7. Fundamental frequency (pitch) contours of the trisyllabic compounds.

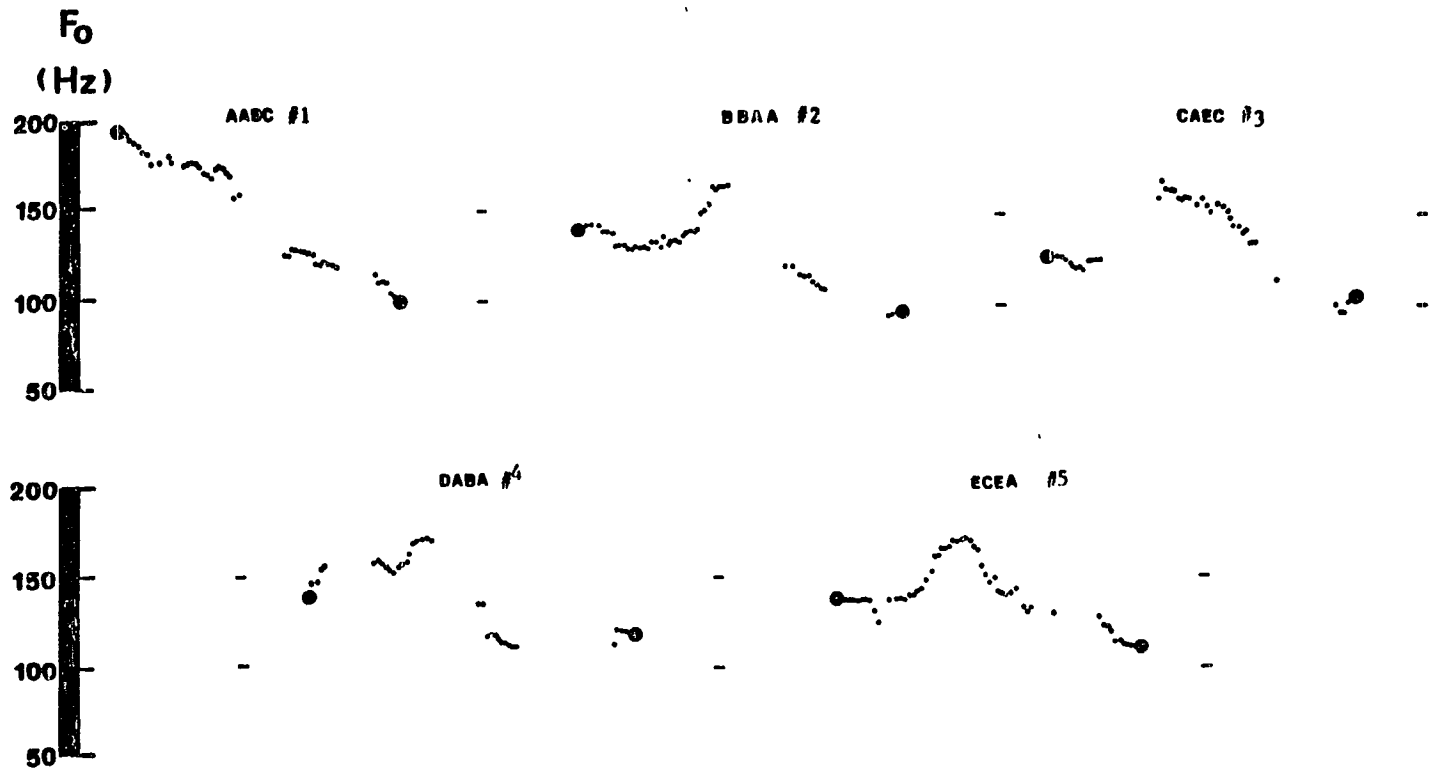


FIGURE 8. Fundamental frequency (pitch) contours of the quadrisyllabic compounds.

4. SUMMARY OF THE RESULTS OF THE ACOUSTICAL ANALYSIS.

1. The shapes of the five etymologically distinct Shanghai 'tones' that occur on monosyllabic words have been analyzed, confirming previous descriptions of the pitch patterns involved.

2. We find there are five main contours that occur on the bisyllabic compounds, plus an additional pattern in the cases where the second syllable is a type D syllable and the first is a type B, C or E syllable.

3. The shape of the pitch contour of a bisyllabic compound is similar to the shape of the pitch contour of its first syllable when it occurs as a monosyllabic word, except for the cases where the second syllable is a type D syllable and the first is a type B, C or E syllable.

4. There are five pitch contours that occur on the trisyllabic compounds.

5. There are only four pitch contours that occur on the quadrisyllabic compounds.

6. The shapes of the pitch contours of the bisyllabic, the trisyllabic and the quadrisyllabic compounds with a type A syllable as their first syllable are all similar to the pitch contour of a type A syllable in isolation.

7. The shapes of the pitch contours of the bisyllabic and the trisyllabic compounds with a type E syllable as their first syllable are similar to the pitch contour of the type E syllable in isolation (given that the final falling portion of the pitch contour of the type E syllable is not considered an intrinsic tonal property).

8. Although there is a difference in the first portion of the pitch contour, the remaining portion is similar in trisyllabic compounds with a type B, C or D syllable as their first syllable.

9. Although there is a difference in the first portion of the pitch contour, the remaining portion is similar in quadrisyllabic compounds with any type of syllable as their first syllable, namely, it falls to low.

10. The first portion of the pitch contours of the trisyllabic and the quadrisyllabic compounds is similar to the beginning points of the pitch contour of the first syllable in isolation.

5. PHONEMIC TONES.

We have discussed the pitch contours found for each of the syllable types A, B, C, D and E in Section 2. We will now present a phonological interpretation of Shanghai tones and tone sandhi.

In monosyllabic words only four contrastive pitch patterns are found. Type A syllables have a high falling pattern, type B a mid rising pattern, types C and E a low rising pattern and type D a high level pattern. Types C and E differ from each other in segmental composition; type C syllables being open or nasal final and long and type E syllables being checked by

a glottal stop and short. We may therefore say that syllables of types C and E do not differ in tone. Although type D syllables are also short and checked, their pitch is not similar to any of the long syllables A, B or C. We need therefore to provide for no more than four different contrastive tone patterns on monosyllabic words.

The question arises whether less than four tonemic contrasts should be recognised in Shanghai phonology. There are two aspects to consider, the relationship between consonants and tones and the possibility of decomposing the contours into more primitive elements. Because voiced obstruents are restricted to syllables of types C and E, it could be argued that the tone pattern of C and E (low rising) is a variant of the same phonological tone(s) as B (mid rising). There would then be a phonological rule lowering the tone after an initial voiced obstruent, and only three underlying tonal patterns on syllables would be recognised (A falling; B, C and E rising; D level). While this rule would provide a partial recapitulation of Shanghai historical tonology there are two principal arguments against its adoption. In the first place, the correlation of consonant voicing and tonal pattern breaks down when sonorants are examined: voiced nasals, laterals and approximants may appear initially in syllables of types A, B and D as well as C and E. Some examples of contrasting sets are given in (5):

(5)	A.	/wə̂ŋ/	"temperature"		A.	/wê/	"fearless'
	B.	/wə̄ŋ/	"stable"		B.	/wḗ/	"to feed'
	C.	/wə̌ŋ/	"faint"		C.	/wê/	"for"
	A.	/mû/	"devil"		D.	/wʌ̍ʔ/	"to pick'
	B.	/mū/	"female"		E.	/wʌ̌ʔ/	"slippery"
	C.	/mǔ/	"to grind"				

Secondly, the voicing of an obstruent is irrelevant in the process of tone sandhi: tones on non-initial syllables change in the same ways regardless of whether these syllables have voiced or voiceless initial obstruents, and no changes in obstruent voicing occur as a result of the tone changes[3]. We therefore maintain that four tonal patterns must be distinguished with types C and E distinguished from type B by some tonal property as well as by initial consonant voicing where necessary. Of course, the important redundancy involved in the relationship of obstruent voicing and tonal category needs to be stated somewhere in the grammar, but it is not a phonological rule which permits a reduction of the number of contrasting tonal patterns. In many tone languages, rising and falling pitch patterns can be shown to be due to the juxtaposition of a sequence of level tones. In other words, a falling contour may be analyzed as High followed by Low and a rising contour is analyzed as Low followed by High. In Shanghai, only type D syllables have a simple level (High) tone and no monosyllabic words with Mid or Low pitch occur. However, we believe there is evidence in favor of analyzing the pitch contours as being sequences of level tones.

3. Sherard (1972) reports that with a following Low tone the voiced stops have a slight breathy (murmured) quality. When the tone is changed the breathiness is absent but the voicing remains.

This evidence is principally to be found in the relationship between the pitch patterns on monosyllabic words and the pitch patterns on bisyllabic compounds and over the first portion of longer compounds. This relationship suggests that the contours should be decomposed and gives an indication of the nature of the underlying elements. As noted in Section 3, the pattern over the first two syllables of a compound is, with some exceptions, similar to that found when its first element occurs in isolation. For example, /di/ 'electricity' has a low rising pitch pattern in isolation; in a compound with /t^{h}i/ 'stairs', the first syllable has a low level pitch and the second syllable a mid level pitch (see #16 on Figure 6). What appears to have happened is that a sequence of 2 tone levels on the first syllable has been split so that each syllable receives one tone. The phonetic sequences of tones on monosyllables might be written in terms of 3 principal levels, High (H), Mid (M) and Low (L) as follows (where M↑ represents a raised Mid level):

SYLLABLE TYPE	PHONETIC TONE(S)
A	[HL]
B	[MM↑]
C, E	[LM↑]
D	[H]

Note however that in the case of the phonetic sequences [MM↑] and [LM↑] there is an alternation with what might be represented as [MH] and [LH] respectively. For example, in /di/ 'electricity' and in /di t^{h}i/ 'escalator' the contour is [LM↑], but in longer compounds with /di/ or another type C syllable as their first element the second syllable of the compound has a high level pitch (see the 3rd row of examples on Figure 7). The rule governing these alternations seems to be that lowering of High takes places when it follows a non-High and precedes a word-boundary. In other words the underlying tones for the monosyllables of the various syllable types are as follows:

SYLLABLE TYPE	PHONEMIC TONE(S)
A	/HL/
B	/MH/
C, E	/LH/
D	/H/

A more formal account of this analysis and further evidence in its favor will be presented below in the detailed statement of sandhi rules.

These rules will be formulated in terms of the revised framework for a suprasegmental analysis of tone proposed by Leben (1978). In an earlier proposal (Leben, 1971b, 1973a, b) Leben assumed that for a language with an underlying suprasegmental level of representation, a phonological rule would map the units in the suprasegmental tone patterns into segmental tone features, and that in phonetic representation tone was expressed as a segmental feature. In the revised version (Leben, 1978), the assumption that one is mapped into a segmental feature has been abandoned, instead Leben accepts Goldsmith's theory of 'autosegmental association between tones and segments" (Goldsmith, 1976a, b) which asserts that the units of suprasegmental tone patterns are not mapped into segmental tone features, instead the tonal representation is associated with the segmental representation at some stage in a phonological derivation but remains suprasegmental at all stages. Before the process of association, tones simply form a property of the word as a whole. After this process, association lines specify which segment(s) or syllable(s) each tone is coarticulated with. Derivations are subject to what Goldsmith has termed the "well-formedness condition":

(6) Well-formedness Condition (WFC):
a. Every tone is associated with some syllable.
b. Every syllable is associated with some tone.
c. Association lines may not cross.

According to Goldsmith,

> 'the "well-formedness condition" is in the indicative not the imperative. A derivation containing a representation that violates the WFC is not thereby marked as ill-formed; rather, the Condition is interpreted so as to change the representation minimally by addition or deletion of association lines so as to meet the Condition maximally.' (Goldsmith, 1976b, p. 27).

This theoretical framework has been chosen because it seems to predict certain processes in Shanghai tone sandhi that would be unexpected if tone was a segmental property.[4] For example, it will be necessary to propose rules that delete tones without deleting the segment(s) with which they are lexically connected. Furthermore there is a process which places a low tone at the end of longer compounds; this tone may be realized over one, two or more syllables. Both of these cases seem to attest to the independence of the tonal and segmental levels in the phonological representation. This question will receive further comment in the course of the formal presentation of the rules.

6.0. TONE SANDHI PATTERNS.

We now turn to the phonological analysis of the tone sandhi patterns

[4]The second author is persuaded of the elegance and simplicity of the autosegmental account of the data but maintains reservations concerning the necessity of positing non-segmental elements in phonology.

observed on compounds. In order to do this, we will first establish a phonetic notation of the sandhi patterns, derived from the acoustical data in Figures 6, 7 and 8. As noted in Sections 3 and 4, the pitch contour of a bisyllabic compound is governed by the underlying tone(s) on the first syllable. In most cases, the pattern found on the first syllable when spoken in isolation is extended to include the second syllable. We therefore represent the patterns on the bisyllables with the same phonetic tones found on monosyllables. Note, however, that type C and type E syllables, although they have the same tones, result in slightly different patterns on bisyllabic compounds of which they form the first element. A type C intial syllable produces the pattern [L M↑], where a type E syllable produces the pattern [L LM↑].

When the second syllable is type D and the first is type B, C or E, the second syllable is higher pitched than when any other type of syllable folows B, C or E. The relationship between the phonetic tones in monosyllables and the bisyllabic compounds is shown in Table V:

	PHONETIC TONE(S) ON MONOSYLLABIC WORDS	PHONETIC TONE(S) ON BISYLLABIC COMPOUNDS WITH RESPECTIVE FIRST ELEMENTS		
				Type D as 2nd element
type A	[HL]	[H L]		
type B	[MM↑]	[M M↑]	~	[M H]
tone C	[LM↑]	[L M↑]	~	[L H]
tone D	[H]	[H H]		
tone E	[LM↑]	[L LM↑]	~	[L H]

TABLE V. Tone patterns on the bisyllabic compounds.

The pitch patterns on the longer compounds are similarly governed by the underlying tone(s) on the first syllable. The first two syllables receive the extended tone pattern of the first element of the compound and the remaining ones are generally low in pitch. A syllable between a high-pitched syllable and a low-pitched syllable will be mid in pitch. As with the bisyllabic compounds, an exceptional pattern occurs if a trisyllabic compound has a type E syllable as its first element. In this case the pattern over the three syllables will be [L L LM↑]. However, quadrisyllabic compounds with a first syllable of type E or type C have the same tone patterns as each other and are not exceptional. Note also that the first of two high-pitched syllables is lower than the second. Table VI presents the phonetic tones of these polysyllabic compounds in relation to the underlying and phonetic tones of the monosyllables occurring as first element.

FIRST ELEMENT OF COMPOUND	UNDERLYING TONES AS MONOSYLLABIC WORDS	PHONETIC TONES AS MONOSYLLABIC WORDS	PHONETIC TONES ON TRI-SYLLABIC COMPOUND WITH RESPECTIVE SYLLABLE TYPES AS FIRST ELEMENT	PHONEMIC TONES ON QUDRI-SYLLABIC COMPOUND WITH RESPECTIVE SYLLABLE TYPES AS FIRST ELEMENT
type A	/HL/	[HL]	[H M L]	[H M L↓ L]
type B	/MH/	[MM↑]	[M H L]	[M H M L]
type C	/LH/	[LM↑]	[L H L]	[L H M L]
type D	/H/	[H]	[H↓ H L]	[H↓ H M L]
type E	/LH/	[LM↑]	[L L LM↑]	[L H M L]

TABLE VI. Tone patterns on the trisyllabic and the quadrisyllabic compounds.

6.1. TONE RULES FOR THE DERIVATION OF THE TONE PATTERNS ON THE COMPOUNDS.

The phonetic tone patterns on the compounds can be derived by a number of simple rules which operate in conjuction with the Well-formedness Condition (WFC). The process of compound formation is viewed as one in which the word boundaries are deleted between the morphemes involved. This may be stated by the informal rule (7).

(7) WORD BOUNDARY DELETION.
When a compound is formed, the internal word boundaries are deleted.

As has already been established, the tone pattern of a compound is determined by the tone(s) on the first syllable only (with one exception). Because the tone(s) on any subsequent syllable make no contribution to the pattern we assume they have been deleted. It can be shown that they are not just subject to an assimilatory change under the influence of the tone(s) on the first syllable, since they are replaced on later syllables by an arbitrary inserted tone. Note that no other features are changed in the course of tone-deletion. Thus, for example,[a] type E syllable retains its initial voicing of an obstruent[4], its short vowel duration and its final glottal stop when it is in second position in a compound (the glottal stop tends to be lost when a third syllable follows).

We will argue that the deletion process does not apply when the second syllable of a bisyllabic compound is a type D syllable (i.e., has a single underlying tone, /H/) and the first syllable has /MH/ or /LH/ underlying tones. Hence the rule describing the process must be formulated with conditions to prevent its application in this environment. This rule is given as (8) where S stands for a syllable and T for the tone or tones associated with that syllable.

(8) TONE DELETION 1.

$$\begin{array}{ccccccccc} \#S_1 & + & S_2 & + & \langle S_3 \rangle & + & \langle S_4 \rangle & + & \ldots\# \\ \| & & | & & \langle | \rangle & & \langle | \rangle & & | \\ \#T_1 & & T_2 & & \langle T_3 \rangle & & \langle T_4 \rangle & & \ldots\# \end{array} \longrightarrow \begin{array}{ccccccccc} \#S_1 & + & S_2 & + & \langle S_3 \rangle & + & \langle S_4 \rangle & + & \ldots\# \\ \| & & | & & \langle | \rangle & & \langle | \rangle & & | \\ \#T_1 & & \emptyset & & \langle \emptyset \rangle & & \langle \emptyset \rangle & & \ldots\# \end{array}$$

Condition: does not apply if

$$T_1 = \begin{Bmatrix} MH \\ LH \end{Bmatrix} \text{ and } T_2 = H \text{ and } S_3 \ldots = \emptyset.$$

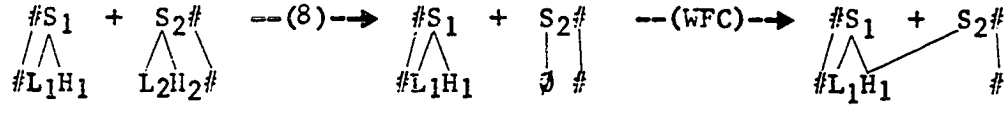
It was suggested in Section 5 that the two rising sequences of tones on monosyllables are underlyingly /MH/ and /LH/. The phonetic output in monosyllables and in most bisyllabic compounds of these tone sequences

contains a lowered High tone, i.e., [MM$^{\uparrow}$] and [LM$^{\uparrow}$]. This argues for a tone lowering rule roughly of the form

(9) H-LOWERING.

$$H \longrightarrow M^{\uparrow} / \begin{Bmatrix} L \\ M \end{Bmatrix} \_\_ \#$$

Apparently the bisyllables with type D syllables in second position are exempt from this rule, as they retain a phonetic [H] tone on the second syllable. It seems explanatory to regard this exemption as due to the fact that these type D syllables retain their original underlying tone and do not receive their tone from the first syllable. Note that the Well-formedness Condition (WFC) will automatically re-associate a tone to the second syllable of bisyllabic compounds following TONE DELETION 1 (i.e., 8), for example:

(10)


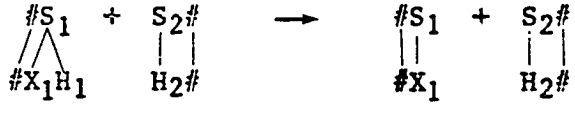
A language specific rule must apply to de-associate the second of two tones on the first syllable, so that only one tone is associated with each syllable. Using the subscript attached to the tones to impose a global condition on the tone-lowering rule, it can be reformulated as (9') to apply to just the monosyllables and those bisyllables which are affected;

(9') H-LOWERING (REVISED).

$$H_1 \longrightarrow M^{\uparrow} / \begin{Bmatrix} L \\ M \end{Bmatrix} \_\_ \#$$

This rule will not apply to bisyllables in which the original H tone remains on the second syllable, i.e., those not affected by TONE DELETION 1 (8). In these cases the tone sequence on the first syllable is simplified by deletion of the second tone. This process is described by the rule in (11):

(11) TONE DELETION 2.



In this rule 'X' is a cover symbol for any tone, although only M or L can occur in this position. Because the tone associated with S_2 is H_2, the HIGH LOWERING (REVISED) (9') does not apply to the output of TONE DELETION 2.

We have discussed in the earlier sections (3-5) how the spreading of the tones from the first syllable is essentially limited to affecting the second syllable of a compound. The tone on a third, fourth or later syllable is the same regardless of the initial syllable (with one exception). Although the tones on these syllables are deleted by TONE DELETION 1 (8), these syllables are not re-associated with the remaining tone(s) of the first syllable. A tone or tones must be inserted in the string instead. We will argue that this insertion rule simply associates a L tone with the third syllable of a compound. In the trisyllables the third syllable is phonetically Low, and in the quadrisyllabies the final syllable is phonetically Low. There is evidence that a Low tone in the environment of H__L is raised to Mid. For example, a compound with a type A syllable as its first element (i.e., with underlying /HL/ on S_1) has the HL sequence spread over the first two syllables. When a third syllable occurs, the tone on the second syllable is raised to Mid. In a quadrisyllabic compound the tone on the third syllable is phonetically Mid when a phonetic High precedes and Low follows. We assume the same tone-raising rule has applied in this instance. Therefore the inserted tone on the third syllable can be taken to be invariably a Low tone. And as the tone on the fourth (and subsequent) syllable is Low on the surface it can be presumed that the Low tone inserted onto the third syllable is spread to any following syllables in the compound. In fact the Well-formedness (WFC) would automatically associate any following syllables (which are not associated with any tone following TONE DELETION 2 (11)) with the inserted tone. The insertion rule applies to all trisyllabic or longer compounds except trisyllabic compounds with a type E syllable their first element, so the rule can be formulated as (12):

(12) L-INSERTION.

$$\begin{array}{c} S_3 \\ | \\ \emptyset \end{array} \longrightarrow \begin{array}{c} S_3 \\ | \\ L \end{array}$$

except if $S_4 = \emptyset$ and $S_1 = (C)\underset{L\ H}{\underset{\wedge}{V}}?$

The exceptional trisyllabic compounds are similar to the bisyllabic compounds with an initial type E syllable in that both of the underlying tones on the first syllable are associated with the final syllable of the compound. This and other matters not automatically accounted for by the WFC are governed by a series of rules which regulate the association between tones and syllables. These rules are as follows:

(13) ASSOCIATION RULES.

(a) Delete the association line between the second of two tones and S_1 in any compound.
(b) Insert an association line between an initial L and the final syllable in a bisyllabic or trisyllabic compound, where S_1 = (C)V?.
(c) Insert an association line between S_2 and the tone to the left when $S_3 \neq \emptyset$ and $S_1 = (C)\underset{\underset{H}{|}}{V}?$.

The operation of these association rules and the previously discussed tone deletion and insertion rules will be illustrated by showing the derivation of some selected compounds.

6.3. DERIVATION OF TONE PATTERNS ON BISYLLABIC COMPOUNDS.

The majority of bisyllabic compounds undergo TONE DELETION 1 (8), and ASSOCIATION RULE (13a). A simple derivation of this kind can be illustrated by a compound which has a type C syllable as its first element and a type A syllable as its second element. The monosyllables /di/ (LH) 'eletricity' and /t^hi/ (HL) ' stairs' combine to form a compound meaning 'elevator' (#11 in Figure 6 and Table 2). Following deletion of the word boundaries between these elements, the condition for TONE DELETION 1 (8) is satisfied, i.e.

(13a) $\#di + t^hi\# \xrightarrow{(8)} \#di + t^hi\#$

$\#L_1H_1 \quad H_2L_2\# \qquad \#L_1H_1 \quad \emptyset\#$

The immediate output of (8) violates the WFC because the second syllable is not associated with any tone, hence an association line is inserted as follows:

(14b) $\#di + t^hi\# \xrightarrow{(WFC)} \#di + t^hi\#$

$\#L_1H_1 \quad \emptyset\# \qquad \#L_1H_1 \quad \#$

The ASSOCIATION RULE (13a) serves to capture the generalization that in compounds of any length the first syllable may be associated with only one tone. It now applies in this derivation, viz.

(14c) $\#di + t^hi\# \xrightarrow{(13a)} \#di + t^hi\#$

$\#L_1H_1 \quad \# \qquad \#L_1H_1 \quad \#$

The sequence $\#L_1H_1\#$ is subject to the H-LOWERING (REVISED)(9'), so that the eventual output from this derivation is as in (14d)

(14d) $\#di + t^hi\# \xrightarrow{(9')} \#di + t^hi\# = [d\grave{i}\ t^h\bar{\acute{i}}]$ 'elevator'

$\#L_1H_1 \qquad \# \qquad\qquad \#L_1M_1 \qquad \#$

Note that there is no need to impose any extrinsic ordering of these rules. They apply whenever their structural description is satisfied.

An additional example of a compound which undergoes the same rules is one with a type B syllable as its first element and other than type D syllable as second element. The example is /pi/ (MH)'flat' + /dɤ/ (LH) 'bean' (#8 in Figure 6 and Table II) which has the derivation in (15).

(15) $\#pi + dɤ\# \xrightarrow{(8)} \#pi + dɤ\#$

$\#M_1H_1 \quad L_2H_2\# \qquad \#M_1H_1 \quad \emptyset\#$

$\xrightarrow{(WFC)} \#pi + dɤ\#$

$\#M_1H_1 \qquad \#$

$\xrightarrow{(13a)} \#pi + dɤ\#$

$\#M_1H_1 \qquad \#$

$\xrightarrow{(9')} \#pi + dɤ\# = [p\bar{i}\ d\bar{\acute{ɤ}}]$ 'flat bean'

$\#M_1M_1 \qquad \#$

When there is a single tone on the first element of a compound there is no work for rule 13a to do. The compound formed from /tcɩ?/ (H) 'urgent' and /cy/ (HL) 'need' (#16 in Figure 6 and Table II) has the derivational history shown in (16).

(16) $\#tcɩ? + cy\# \xrightarrow{(8)} \#tcɩ? + cy\#$

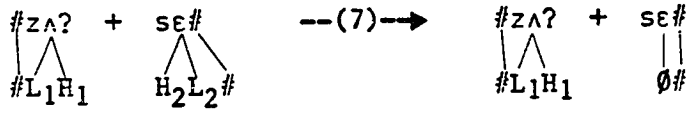
$\#H_1 \quad H_2L_2\# \qquad \#H_1 \quad \emptyset\#$

$\xrightarrow{(WFC)} \#tcɩ? + cy\#$

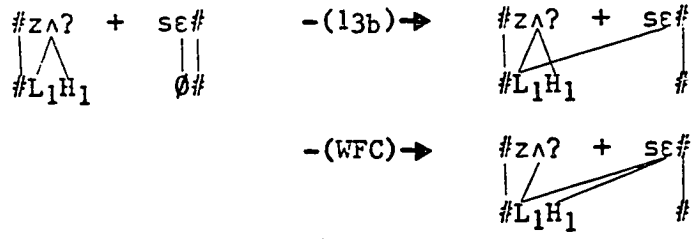
$\#H_1 \qquad \#$

The final output is [tcí cý] 'urgent need' where the usual deletion of a medial glottal stop has also occurred.

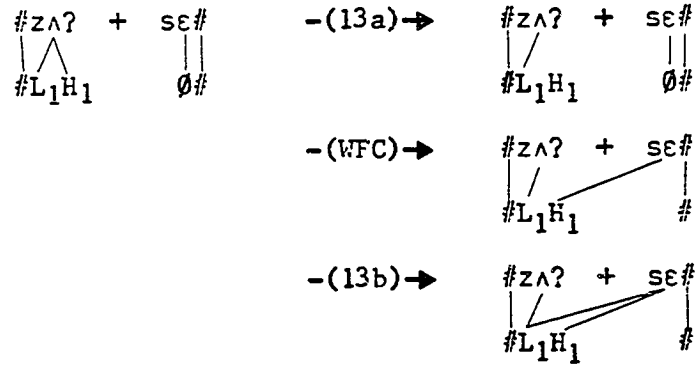
Where the first element of a bisyllabic compound is type E, TONE DELETION 1 (8) applies as with (14-16). However, ASSOCIATION RULE (13b) applies. An example is /zʌ?/ (LH) 'ten' + /sɛ/ (HL) 'three' (#21 in Figure 6 and Table II). The tones on the 2nd syllable are deleted:

(17a) 

The ASSOCIATION RULE (13b) directs that the initial L be associated with the final syllable. When the required association line is inserted, the WFC intervenes to ensure that association lines do not cross. The minimal change which will satisfy the WFC is to associate the H tone to the final syllable and delete the association line between H and the first syllable. Thus,

(17b) 

It is possible that (13a) applies before (13b) in this case. We would then have (17c) instead of (17b).

(17c) 

The output of these derivations is the same and there seems no need to choose between them. To do so would require abandoning the simple condition of rule application that rules apply whenever their structural description is met. In any order of application the rules will produce the final output [zʌ̀ sɛ̄] 'thirteen' via the H-LOWERING (REVISED)(9') and the optional glottal stop deletion rule.

6.4. DERIVATION OF TONE PATTERNS ON LONGER COMPOUNDS.

The longer compounds are affected by the L-INSERTION rule (12) as well as WORD BOUNDARY DELETION (7), the TONE DELETION 1 and 2 (8, 10) and the ASSOCIATION RULES (13). The most frequent kind of derivation can be exemplified by a trisyllabic compound which has a type B syllable as its first element, for example a compound formed from /jiɤ/ (MH) 'young' + /zɽ/ (LH) 'juvenile' + /yø/ (LH) 'garden' (#8 in Figure 7 and Table III). Following deletion of the internal word boundaries, TONE DELETION 1 (8) applies:

(18a) #jiɤ + zɽ + yø# --(8)--> #jiɤ + zɽ + yø#

#M$_1$H$_1$ L$_2$H$_2$ L$_3$H$_3$# #M$_1$H$_1$ ø øø#

As there is a third syllable in this compound a Low tone is inserted into the tone string by L-INSERTION (12):

(18b) #jiɤ + zɽ + yø# -(12)--> #jiɤ + zɽ + yø#

#M$_1$H$_1$ ø øø# #M$_1$H$_1$ ø L$_3$#

Of course, the output of (12) in (18b) as represented here violates the WFC since the second syllable is not associated with any tone. However, ASSOCIATION RULE (13a) directs that the association line between S_1 and the second of two tones be deleted. This produces the output in (18c):

(18c) #jiɤ + zɽ + yø#

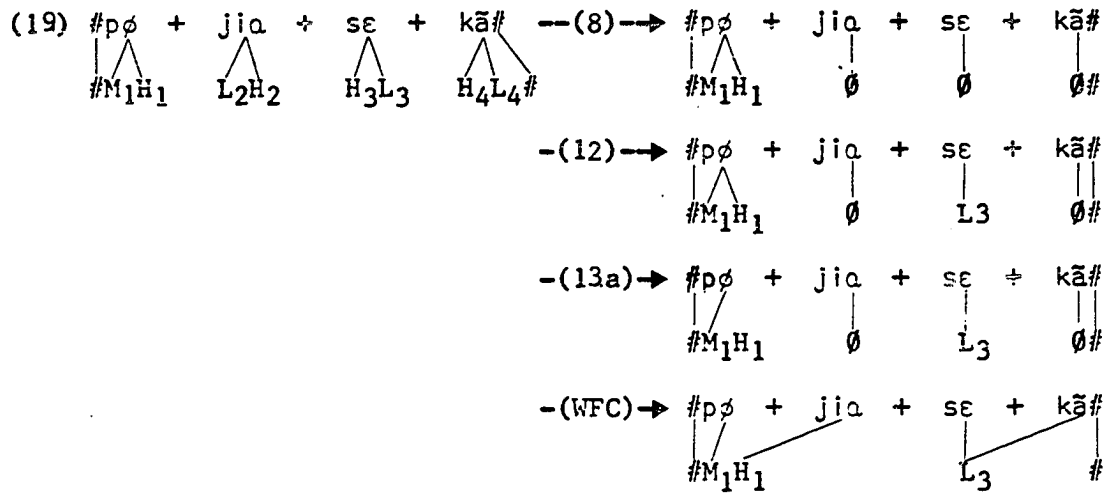
#M$_1$H$_1$ ø L$_3$#

The WFC intervenes to correct the violation by making the minimal change, that is, an association line between the unassociated second tone and the second syllable is inserted. The output is (18d) with tone pattern [M H L].

(18d) #jiɤ + zɽ + yø# = [jiɤ̄ zɽ́ yø̀] 'kindergarten'

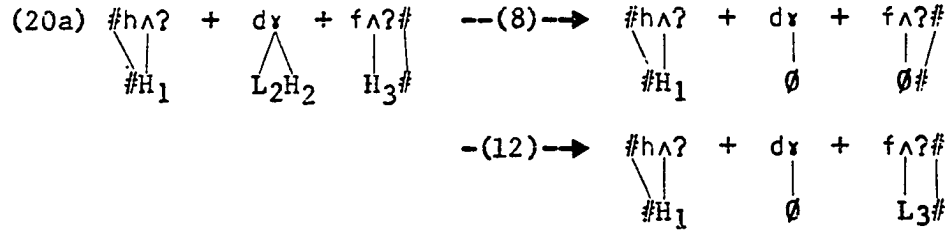
#M$_1$H$_1$ L$_3$#

The derivation of a quadrisyllabic compound with an initial syllable of the same type is essentially similar, except that the WFC associates the inserted Low tone on S_3 to the fourth syllable as well. An example is the compound formed by combining /pø/ (MH) 'half' + /jiɑ/ (LH) 'night' + /sɛ/ (HL) 'three) + /kã/ (HL) 'time measurement' (#2 in Figure 8 and Table IV) whose derivation is given in (19).

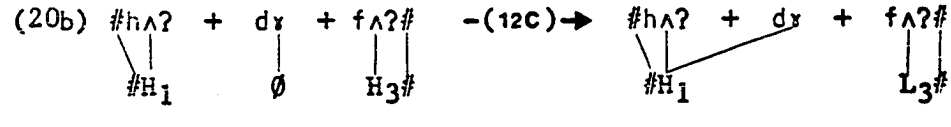


The eventual phonetic output from this derivation shows the effect of the tones on each other. Because of co-articulation effects between H on the second syliable to the following L, the pitch of the third syllable is closer to a phonetic Mid level and the compound is transcribed [pø̄ jiá sɛ̄ kã] 'pre-dawn'. As this kind of coarticulation between adjacent tones is a straightforward phonetic effect it is not discussed further in this presentation and no rule has been formulated.

In the derivation of longer compounds with an initial type D syllable the ASSOCIATION RULE (13c) applies. For example, the trisyllabic compound formed from /hʌ?/ (H) 'black' + /dɤ/ (LH) 'head' + /fʌ?/ (H) 'hair' (#18 in Figure 7 and Table III) has the derivation in (20).

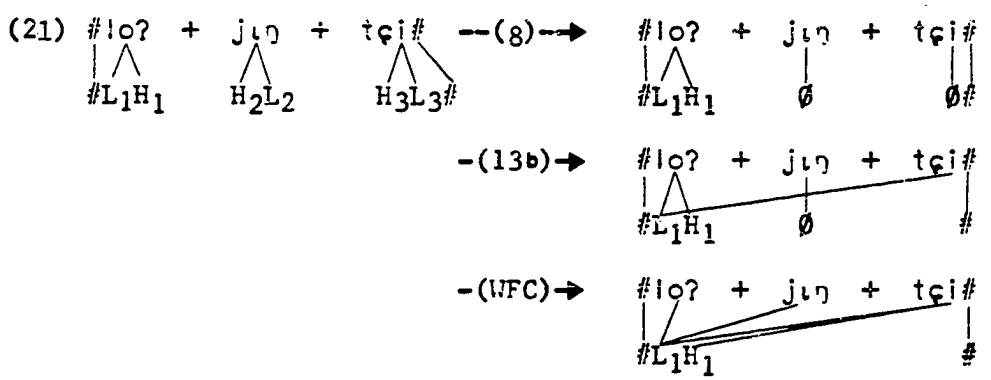


The output of (12) violates the WFC because the second syllable is not associated with any tone. However, the WFC alone does not uniquely resolve this case since S_2 might be associated with either the tone to the left or the tone to the right. The correct output must associate the second syllable with the H tone on S_1 and (13c) inserts an association line which does so.



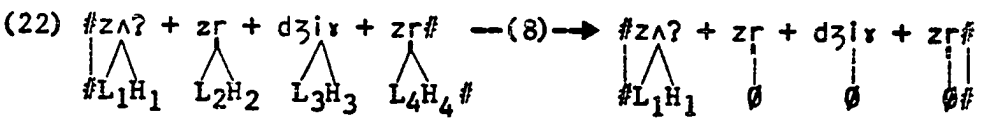
There seems to be a slight lowering of the pitch of the first of two syllables associated with an initial H tone when a third syllable follows in the compound. This lowered high pitch is represented in Table IV by H↓ and the final phonetic output from this derivation is [hʌ́↓ dɪ́ fʌ̀ʔ] 'black hair'. This phonetic lowering rule applies also to the quadrisyllabic compounds with an initial type D syllable, see #4 in Figure 8 and Table IV.

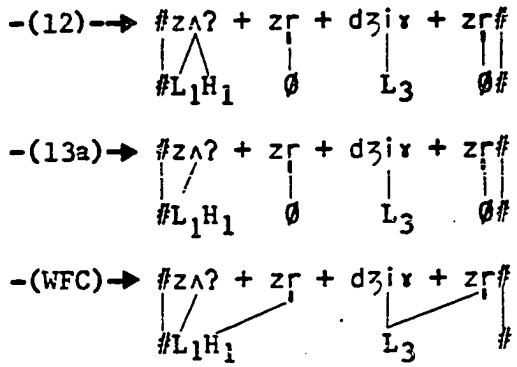
The longer compounds with an initial type E syllable undergo different derivations according to the number of syllables they contain. Trisyllables are affected by rule (13b), as are bisyllables with an initial type E syllable, but quadrisyllabic compounds share the same kind of derivation as most of other longer compounds exemplified in (19). A trisyllabic compound of this type is formed from /lo?/ (LH) 'to record' + /jɪŋ/ (HL) 'sound' + /tɕi/ (HL) 'machine' (#21 in Figure 7 and Table III) and its derivation is given in (21). Note that L-INSERTION (12) does not apply in this derivation.



As with the derivation of the bisyllabic compound in (17) the WFC intervenes to correct the violation, namely crossed association lines, that would otherwise result from the application of (13b). Again as with (17), it is immaterial if (13a) does nor does not apply to this case. In either case the eventual output, including the effect of H-LOWERING (REVISED)(9'), is [lò jɪ̀ŋ tɕī] 'tape recorder'.

The difference in the derivation of a quadrisyllabic compound which has a type E initial syllable can be illustrated with the compound formed from /zʌʔ/ (LH) 'factual' + /zr̩/ (LH) 'to be' + /dʒiɹ/ (LH) 'then' + /zr̩/ (LH) 'to be'. Compare the derivation in (22) with that in (19). The derivation in (22) can be seen to be like the normal quadrisyllabic compound illustrated in (19).





The eventual phonetic output is [zʌ̀ zɹ̩́ dʒiɤ̄ zɹ̩̀] 'unpretentious' where the co-articulation between adjacent tones is noted by representing the pitch of the third syllable as in the mid range. Any other compounds undergo similar derivations.

7. CONCLUSION.

We now return to the question of the accuracy of previous statements in the linguistic literature on Shanghai tone sandhi. The regular patterns of sandhi we have found show two major processes, the spreading of the tone(s) of the first syllable of a compound to the second syllable, and the replacement of the tone(s) on third and subsequent syllables by Low tone. There are a few minor patterns in addition, but, in the majority of patterns, only the tone of the first syllable matters in the output.

Sherard's observations on tone sandhi do not agree entirely with ours. A significant difference between the data in this paper and the contours reported by Sherard concerns the longer compounds. Sherard reports that for compounds with type B or D initial syllables the contour for the compounds is same as that on the initial syllables when spoken as monosyllables (i.e., mid-dipping and high level, respectively). Sherard's data on compounds with type, C or E syllables agree in the main with ours. It seems to us that the differences arise because of two factors. Principally, as Sherard himself notes, the linguistic situation in Shanghai has been much influenced by immigration from surrounding areas. In particular, the speech variety that Sherard records seems to contain features characteristic of the dialect of Ningpo[5], to the south of Shanghai. In Ningpo, disyllabic compounds with type B and D initials do have the contours that Sherard reports. No Low tone insertion applies to these cases. For example, Sherard cites the compound formed from /ʌ?/ (H) 'to suppress' + /sə/ (MH) 'age' + /di/ (LH) 'money', referring to money given at the New Year, as [ʌ́ sə́ dí] with a High tone throughout, rather than as [ʌ́ sə́ dì] with final Low, which is the form that would occur in our sub-

[5]Ningpo is a town about 100 miles south of Shanghai. The first author is personally acquainted with the speech of this area through friends and relations who originate from there.

ject's speech. In some other cases it appears that Sherard may be reporting a form that was not read as a true compound. Our overall conclusion is that our data cannot necessarily be interpreted as correcting Sherard where they differ. Our data is more complete than Sherard's but apparently refers to a slightly different dialect. We would claim that our subject's speech is more truly representative of the indigenous Shanghai dialect whose historical connections have been more with the other Northern Wu dialects surrounding the city until recent times.

Ballard (1977) claimed that tone sandhi in Shanghai consists only of a rightward spreading (or in our terms a rightward reassociation) of the tone(s) of the initial syllable. This claim is more justified on the basis of the data in Sherard than it would be for ours but even so it does not account for all the patterns that Sherard reports. For example a compound beginning with a type C syllable (with a low-mid rise in isolation) has a pattern which "starts from a low level, rises to a moderately high pitch and then begins to fall back toward low level" (Sherard, 1972: 102). We recognise our High-Lowering (9') and Low-Insertion (12) rules in relating the tones on the monosyllable and the compound here. Such examples show that there is more than rightward spreading involved even in the sub-dialect Sherard examined. In particular, there is no case for regarding the sandhi process as one in which a contour is extended as a unit over the compound. With the majority of syllables in Shanghai having two underlying tones, it is natural that it is precisely the third syllable of a compound which receives an inserted Low tone, and while this single Low tone may spread over (be associated with) several syllables, contours as such do not spread.

REFERENCES

Anderson, S. R. (1978) "Tone features" Tone: A Linguistic Survey (V. A. Fromkin, ed.) New York, Academic Press: 133-175.

Ballard, W. L. (1977) On some Aspects of Tone Sandhi. Atlanta, Georgia State University.

Goldsmith, J. (1976a) "An overview of autosegmental phonology." Linguistic Analysis, 2, 23-68.

Goldsmith, J. (1976b) Autosegmental Phonology. Ph.D. Dissertation (MIT). Repr. by Indiana University Linguistics Club.

Hangyu Fangyan Gaiyao (A precis of Chinese dialects, 1960). Peking, Wenzi Gaige Chubanshe.

Hombert, J-M (1978) "Consonant types, vowel quality, and tone" Tone: A Linguistic Survey (V. A. Fromkin, ed.) New York, Academic Press: 77-111.

Jiangsushen He Shanghaishi Fangyan Gaikuang (Synopsis of Shanghai dialect and the dialects in Jiangsu Province, 1960). Nanking, Jiangsu Renmin Chubanshe.

Kennedy, G. A. (1953) "Two tone patterns in Tangsic" Language 29: 367-373.

Leben, W. R. (1971) "Suprasegmental and segmental representation of tone." Studies in African Linguistics, Supp. 2, 183-200.

Leben, W. R. (1973a) "The role of tone in segmental phonology." Consonant Types and tone (L. Hyman, ed.). Los Angeles, University of Southern California, Linguistics Program: 115-149.

Leben, W. R. (1973b) Suprasegmental Phonology. Ph.D. Dissertation (MIT).

Leben, W.R. (1978) "The representation of tone" Tone: A Linguistic Survey (V. A. Fromkin, ed.). New York, Academic Press: 177-219.

Ohala, J and Ewan, W. (1972) "Speed of pitch change." Paper presented at the 94th ASA Meeting, Autumn, 1972, Miami Beach, Florida.

Sokolov, M. V. (1965) "An experimental investigation of the tones in the Shanghai dialect." Phonetica, XII, 197-200.

Sherard, M. (1972) Shanghai Phonology. Ph.D. Dissertation (Cornell University).

Traill, A. (in progress) Phonetic and Phonological Sketches of !xóõ Bushman. Ph.D. Dissertation (University of Witwatersrand).

Section 7: *A Spectrographic Investigation of Mandarin Tone Sandhi*

INTRODUCTION

> "In Mandarin there are four tones for stressed syllables. If the average range of the speaker's voice is divided into four equal intervals separated by five points: 1 low, 2 half-low, 3 middle, 4 half-high, and 5 high, any tone can be fiarly well presented by giving its starting and ending pitch, and in the case of circumflex tones, the turning point. Moreover, if we use a short vertical line as a reference line for ordinates and plot a simplified graph to its left, with time as abscissa and pitch as ordinate, we get a letter-like symbol to represent the tone, as in the fast column of the following table:

Tone	Chinese name	Description	Pitch	Graph
1st	Inpyng	high-level	55:	¬
2nd	Yangpyng	high-rising	35:	˥˦
3rd	Shaang	low-dipping	214:	˨˩˦
4th	Chiuh	high-falling	51:	˥˩

> ... Tone sandhi is the change in the actual value of tones when syllables are spoken in succession.
>
> ... When a syllable is completely unstressed, its tone disappears and is said to be atonic or in the neutral tone." (Chao, 1948, p. 24-25, 27)

This paper is a spectrographic investigation of the tone sandhi rules in Mandarin Chinese. It consists of three parts, each dealing with one of the following claims with regard to the tonal phenomena in the language:

CLAIM I:

In a bisyllabic compound, tone 3 /214/ changes to tone 2 [35] in normal speech when it is followed by another tone 3 /214/ (Chao, 1948), that is,

$$/214 + 214/ \rightarrow [35 + 214] \quad \text{(Rule 1)}$$

CLAIM II:

In a trisyllabic compound, tone 2 /35/ on the second syllable changes to tone 1 [55] for speech at conversational speed, when it is preceded by another tone 2 /35/ or tone 1 /55/ and followed by

any except the neutral tone (Chao, 1948, 1968, Cheng, 1968), i.e.,

$$/ \begin{Bmatrix} 55 \\ 35 \end{Bmatrix} + 35 \ X/ \rightarrow [\begin{Bmatrix} 55 \\ 35 \end{Bmatrix} + 55 + X] \quad \text{(Rule 2)}$$

$$X \neq \text{neutral tone}$$

CLAIM III:

In rapid speech, if the tones on all syllables are tone 3 /214/ in a sentence such as,

/lau li mai çiau pi/ (214 214 214 214 214)
'Old Li buys small pen'

then,

(a) tone 3 on the first syllable changes to tone 2 [35],
(b) tone 3 on the second, third and fourth syllables change to tone 1 [55], and
(c) tone 3 on the last syllable remains unchanged (Cheng, 1968),

that is,

/214 + 214 + 214 + 214 + 214/ → [35 + 55 + 55 + 55 + 214] (Rule 3)

In the following sections, we will investigate whether these claims are valid. As the claims were based on the result of impressionistic analysis, they may not accurately describe the actual changes in the fundamental frequency contour when the syllables are juxtaposed. Spectrographic analysis demonstrates physically the actual changes in the fundamental frequency contours for the words and sentence concerned. Thus, the purpose of our study is to provide acoustical data relevant for the analysis of tonal phenomena we have just described.

Two female native Peking speakers who recently came to study in this country participated in our investigation. They, both in the late twenties, were born and grew up in Peking. Their speech is representative of the style of the younger generation in Peking as judged by other native Peking speakers who also arrived in this country recently. Spectrograms were made from their recorded speech. Fundamental frequency measurements were made on the spectrograms.

SECTION (I)

In order to determine whether tone 3 /214/ actually becomes tone 2 [35] when it is followed by another tone 3 /214/, we may compare the fundamental frequency contours of the following two sequences of tones:

(a) /214 + 214/
(b) / 35 + 214/

As tone 2 /35/ is not reported to change in /35 + 214/, the F_0 contours of these two sequences of tones should be the same. Five pairs of bisyllabic compounds were chosen on the basis that they were commonly used. One of the paired bisyllabic compounds had the tone sequences /214 + 214/, and the other /35 + 214/. The five pairs of bisyllabic compounds are shown below:

/tʰu kaɪ/	(214 214)	'land re-distribution'
/tʰu kaɪ/	(35 214)	'to retouch'
/tʂau xuo/	(214 214)	'to look for fire'
/tʂau xuo/	(35 214)	'on fire'
/tçʰi ma/	(214 214)	'at least'
/tçʰi ma/	(35 214)	'to ride a horse'
/mai ma/	(214 214)	'to buy a horse'
/mai ma/	(35 214)	'to bury a horse'
/fən tʂʰaŋ/	(214 214)	'a flour factory'
/fən tʂʰaŋ/	(35 214)	'graveyard'

In order to avoid speakers' conscious effort to contrast the paired bisyllabic compounds, two separate reading lists were made for two separate recording sessions. In one reading list only the bisyllabic compounds with the tone sequence /214 + 214/ were included, and in the other only those with the tone sequence /35 + 214/. In each reading list other dummy meaningful bisyllabic compounds were added to avoid monotony. The test compounds were arranged in a random order. Each test compound was repeated four times in the reading list, and it was placed in the carrier frame as below:

[uo çian-tʂaɪ tu ___ gei ni tʰin]
I now read for you listen

Speakers were instructed to read the word list slightly faster than the normal rate of speech. The recordings were made either in a sound treated booth or in a quiet room.

Forty spectrograms of the bisyllabic compounds (1 pair x 5 bisyllabic compounds x 4 repetitions) were made for the speech of each speaker. They are shown in Figures 1a-1e for Speaker Q.M. and Figures 2a-2e for Speaker Y.H.J. (spectrogram size 78% reduced from the original). Each figure shows 4 repetitions of each of the paired bisyllabic compounds. In each figure, the upper four spectrograms are the bisyllabic compounds with the tone sequence /35 +214/, and the lower four those with tone sequence /214 + 214/. The beginning and end points of the harmonics were

marked with two arrows. Tables Ia, Ib and Ic show the F0 values for the beginning, dip (when there is a dip) and end points of the pitch contour of the first syllable in the paired bisyllabic compounds with the tone sequences /214 + 214/ and /35 + 214/ for both speakers. The F_0 values were obtained from the spectrograms shown in Figures 1a-1e (Spkr. Q.M.) and Figures 2a-2e (Spkr. Y.H.J.).

The shapes of the F_0 contours (see Figure 1a for Spkr. Q.M. and Figure 2a for Spkr. Y.H.J.) of the first syllables in the bisyllabic compound /tʰu kaɪ/ (35 + 214) 'to retouch' (the upper four) and those of the first syllables in the bisyllabic compounds with the /214 + 214/ tone sequence meaning 'land re-distribution' (the lower four) are quite similar and they are both rising. This shows that the tone on the first syllable in /tʰu kaɪ/ (214 + 214) 'land re-distribution' has indeed changed from a dipping /214/ to a rising contour for both speakers. This is also true for the bisyllabic compound /tɕʰi ma/ (214 + 214) 'at least' for both speakers as shown in Figures 1c (Spkr. Q.M.) and 2c (Spkr. Y.H.J.). Notice however, the F_0 values for the dip and end points of the pitch contour are higher for the first syllable in the bisyllabic compound with the tone sequence /35 + 214/ than for the first syllable in the bisyllabic compound with the tone sequence /214 + 214/, for example, as shown in Table Ia, the average F_0 values for the dip and end points are 213.8 Hz and 316.6 Hz respectively for [tʂau] in /tʂau xuo/ (35 + 214) and 199.9 Hz and 244.4 Hz respectively for [tʂau] in /tʂau xuo/ (214 + 214) for Speaker Q.M., and correspondingly 202.0 Hz/293.5 Hz and 182.1/271.4 Hz for Speaker Y.H.J.

The shapes of the F_0 contours of the first syllables in the bisyllabic compounds with the /35 + 214/ tone sequence (the upper four) and those of the first syllables in the bisyllabic compounds with the tone sequence /214 + 214/ in Figures 1b, 1d and 1e for Speaker Q.M. and Figures 2b, 2d and 2e for Speaker Y.H.J. are also similar in all cases. However, the tone on the first syllable in the bisyllabic words with the tone sequence /214 + 214/ does not change to a rising contour but retains its lexical tone shape /214/. On the other hand, the tone on the first syllable in the bisyllabic compounds with the tone sequence /35 + 214/ has changed from /35/ to a dipping contour. Notice that the F_0 value for the beginning point of the pitch contour is not always higher for the bisyllabic compound with the tone sequence /35 + 214/, for example, as shown in Table Ia, the average F_0 value is 240.5 Hz for [tʰu] in /tʰu kaɪ/ (35 + 214) and 240.9 Hz for [tʰu] in /tʰu kaɪ/ (214 + 214) for Speaker Q.M., and correspondingly 229.0 Hz and 232.6 Hz for Speaker Y.H.J.

Table Id presents the durations (in seconds) for each of the bisyllabic compounds for both speakers. We can see that the duration of the bisyllabic compounds never exceeds 0.57 sec for Speaker Q.M. and 0.71 sec for Speaker Y.H.J., which shows that these bisyllabic compounds were produced at a fairly fast speech rate by both speakers.

To summarize:

(1) the shapes of the F_0 contours of the first syllables in the paired bisyllabic compounds are similar.
(2) The tone on the first syllable in some bisyllabic compounds with the tone sequence /214 + 214/ changes to a rising contour and in others it remains unchanged.
(3) The tone on the first syllable in some bisyllabic compounds with the tone sequence /35 + 214/ changes to a dipping contour and in others it remains unchanged.
(4) The level of F_0 for the dip and end points of the F_0 contour is higher for the first syllable in the bisyllabic compounds with the tone sequence /35 +214/ than for the first syllable in the bisyllabic compounds with the tone sequence /214 + 214/.
(5) The level of F_0 for the beginning point of the F_0 contour is not always higher for the first syllable in the bisyllabic compounds with the tone sequence /35 + 214/.
(6) The above (1)-(5) are true for both speakers.

CLAIM I above states:

/214 + 214/ → [35 + 214] (Rule 1)

According to our acoustical data, the tone /214/ on the first syllable does not always change to a rising contour, that is, it sometimes retains its lexical tone shape. The only time the tone /214/ on the first syllable changes to a rising contour is when the vowel in the first syllable is preceded by aspiration. Furthermore, as we have shown, when it does change to a rising contour, the end point is not as high as it is when the first syllable is underlying tone 2 /35/. Thus, CLAIM I (or Rule 1) does not apply to the data provided by our speakers. On the other hand, tone 2 /35/ on the first syllable changes to a dipping contour when the vowel or diphthong in the first syllable is preceded by aspiration. Based on the observation, we formulate the following rules:

$$/35/ \rightarrow [215] \;/\; \underset{[-Asp]}{C\_\_\_} \;/214/ \quad \text{(Rule 4)}$$

$$/214/ \rightarrow [34] \;/\; \underset{[+Asp]}{C\_\_\_} \;/214/ \quad \text{(Rule 5)}$$

Rule (4) and Rule (5) correctly describe the data provided by the two native speakers of Peking. Tone change in the bisyllabic compounds with the tone sequences /35 + 214/ and /214 + 214/ in Mandarin Chinese is less straightforward than what was previously claimed to be.

SECTION II.

In this section we will investigate whether the F_0 contour of the 2nd-tone /35/ on the second syllable changes to [55] for speech at conversational speed when it is preceded by another 2nd-tone /35/ or a 1st-tone /55/ and followed by any except the neutral tone (CLAIM II), or,

$$/\begin{Bmatrix}55\\35\end{Bmatrix} + 35 + X/ \rightarrow [\begin{Bmatrix}55\\35\end{Bmatrix} + 55 + X] \quad \text{(RULE 2)}$$

$$X \neq \text{neutral tone}$$

The following trisyllabic compounds were used for investigation:

/tuŋ nan fən/	(55 35 55)	'Southeast wind'	from Cheng 1968
/çian ʐən tçaŋ/	(55 35 214)	'cactus'	from Cheng 1968
/mei ian faŋ/	(35 35 55)	'(a name)'	from Chao, 1968
/çi iaŋ şən/	(55 35 55)	'American ginseng'	from Chao, 1968
/san nian tçi/	(55 35 35)	'third grade'	from Chao, 1968
/tsʰuŋ iou piŋ/	(55 35 214)	'onion oil cake'	from Chao, 1968
/tuŋ xɤ ian/	(55 35 51)	'East Riverside'	from Chao, 1968
/fən şuei liŋ/	(55 35 214)	'watershed'	from Chao, 1968
/şuei naŋ fei/	(35 35 55)	'Who can fly?'	from Chao, 1968
/xai mei uan/	(35 35 35)	'not yet finished'	from Chao, 1968
/xan şu piau/	(35 35 214)	'thermometer'	from Chao, 1968

The two female native speakers of Peking described in Section (I) read the above words at a fast speech rate. The recording was made in a sound-treated booth or a quiet room. Spectrograms of each of the above test trisyllabic compounds were made for both speakers as shown in Figures 3.1-3.11 (original size, unreduced), and F_0 values were calculated for the beginning and end and, where relevant, for the central portion of a dipping contour. Table II shows the F_0 values obtained for Speaker Q.M. The trisyllabic compounds listed in Table II are arranged according to the shape of the F_0 contour on the second syllables as indicated in the last column of the table. We can see that tone 2 /35/ on the second syllable does not always change to [55], that is, HIGH LEVEL. In fact, in most cases (9 out of 11) the tone does not. It may change to a MID LEVEL [33] (approximately) as illustrated by compounds #3 and #4 (see also Figures 3.3 and 3.4), to a HIGH DIPPING [535] (approx.) as illustrated by compound #5 (see also Figure 3.5), or it may remain unchanged, as illustrated by compounds #6-11 (see also Figures 3.6-3.11).

Table III shows the F_0 measurements of the trisyllabic compounds provided by Speaker Y.H.J. The spectrograms of the trisyllabic compounds are shown in Figures 4.1-4.11. Tone 2 /35/ on the second syllable in this case changes to even more different shapes. It changes to HIGH LEVEL [55] (approx.), as illustrated by compound #1 (see also Figure 4.1), to MID LEVEL [33] (approx.), as illustrated by compound #2 and #3 (see also Figures 4.2 and 4.3), to FALLING [31] (approx.), as illustrated by compounds #4 and #5 (see also Figures 4.4 and 4.5), and to DIPPING [313] (approx.), as illustrated by compounds #8-#11 (see also Figures 4.8-4.11), or it may remain unchanged, that is, RISING [35] (approx.), as illustrated by compounds #6 and #7 (see also Figures 4.6 and 4.7). Notice that for Speaker Y.H.J. there is only one out of eleven cases in which tone 2 /35/ changes to HIGH LEVEL [55].

Our findings based on the data provided by the two native Peking speakers apparently do not conform to CLAIM II (thus Rule 2). The discrepency between our acoustical data and the earlier impressionistic studies may be due to the fact that the F_0 contour on the second syllable is difficult to perceive because of its short duraton. Table IV shows the durations (in seconds) for the first and the second syllables in each of the trisyllabic compounds. We can see that for Speaker Q.M. the total duration for the two syllables in any trisyllabic compounds does not exceed 0.38 sec and for Seeaker Y.H.J. 0.43 sec. As the duration is so short, we can understand why the changing F_0 on the second syllable was not perceived by the authors of the earlier studies. To conclude, we have demonstrated that CLAIM II is not a valid generalization at the productive and acoustic levels. It may remain valid as an observation of perceived tone changes.

SECTION (3)

CLAIM III (Cheng, 1968) is concerned with tone change at the sentence level in Mandarin Chinese. In rapid speech (faster than fast speech, according to Cheng, 1968), if the tones on all the syllables in a sentence are 3rd-tones /214/, then, (a) the 3rd-tone on the first syllable changes to a 2nd-tone [35], (b) the 3rd-tone on the second, third and fourth syllables change to a 1st-tone [55], and (c) the 3rd-tone on the last syllable remains unchanged, or,

/214 + 214 + 214 + 214 + 214/ → [35 + 55 + 55 + 55 + 214] (RULE 3)

In this section, we will demonstrate whether CLAIM III is phonetically valid. The same sentence given in Cheng (1968) was used in our investigation and it is repeated in the following:

/lau li mai çiau pi/ (214 + 214 + 214 + 214 + 214)
'Old Li buys small pen'

The two female native speakers of Peking were instructed to read the sentence first in fast speech and then in the fastest speed they possibly could. Spectrograms of the sentence produced with the two different speeds are shown in Figures 5a (fast speech) and Figure 5b (fastest speech) for Speaker Q.M. and Figure 6a (fast) and Figure 6b (fastest) for Speaker Y.H.J.

Table V shows the F_0 values for the beginning and end points of the F_0 contour on each one of the first four syllables [ɪau] [ɪɪ] [maɪ] [çɪau] in the sentence produced at these two different speeds by Speakers Q.M. and Y.H.J. These values were obtained from the spectrograms shown in Figures 5a, 5b, 6a and 6b. Also shown are the difference between these F_0 values in each syllable and the total duration of the four syllables produced with two speeds for both speakers. Table VI, which is based on the values shown in Table V, shows the shape of the F_0 contour on each one of the first four syllables for Speakers Q.M. and Y.H.J. As far as the tone on the first syllable [ɪau] is concerned, it has indeed changed from Dipping /214/ to a RISING contour [35] in both speeds and for both speakers. However, the shapes of the F_0 contours on the other syllables [ɪɪ], [maɪ], [çɪau] are far from being level, except for the cases of [çɪau] in the fastest speech for Speaker Q.M. and [ɪɪ] in fast speech for Speaker Y.H.J. They are either RISING or FALLING. Furthermore, Speakers Q.M. and Y.H.J. have produced opposite results for the syllable [maɪ]. The shapes of the F_0 contours on the syllable are RISING in both speeds for Speaker Q.M., but FALLING in both speeds for Speaker Y.H.J.

We have demonstrated that in most cases the tones /214/ on the second third and fourth syllables /ɪɪ/, /maɪ/, /çɪau/ do not change to HIGH LEVEL [55] as Cheng (1968) has claimed that they should, although the tone on the first syllable in the sentence did change to RISING in all cases. Thus, CLAIM III is erroneous insofar as the tone change on the second, third and fourth syllables is concerned. The difference between our results and CLAIM III may be again attributed to the short durations of the syllables. As shown in Table V, the total time of the four syllables [ɪau ɪɪ maɪ çɪau] is only 0.83 sec (fast speed) or 0.55 sec (fastest speed) for Speaker Q.M. and 0.79 (fast speed) or 0.53 sec (fastest speed) for Speaker Y.H.J.

CONCLUSION

Results of the spectrographic analysis of the data provided by two female native speakers of Peking have shown that none of the three claims proposed by Chao, 1948, 1968 and Cheng, 1968 applies to tone sandhi at the productive or acoustic level in the speech of today's young generation of Peking Mandarin speakers.

REFERENCES

Chao, Y.R. (1948). Mandarin Primer. Cambridge, Harvard University Press.

Chao, Y.R. (1968). A Grammar of Spoken Chinese. Berkeley, University of California Press.

Cheng, C.C. (1968). Mandarin Phonology. Ph.D. Dissertation (University of Illinois).

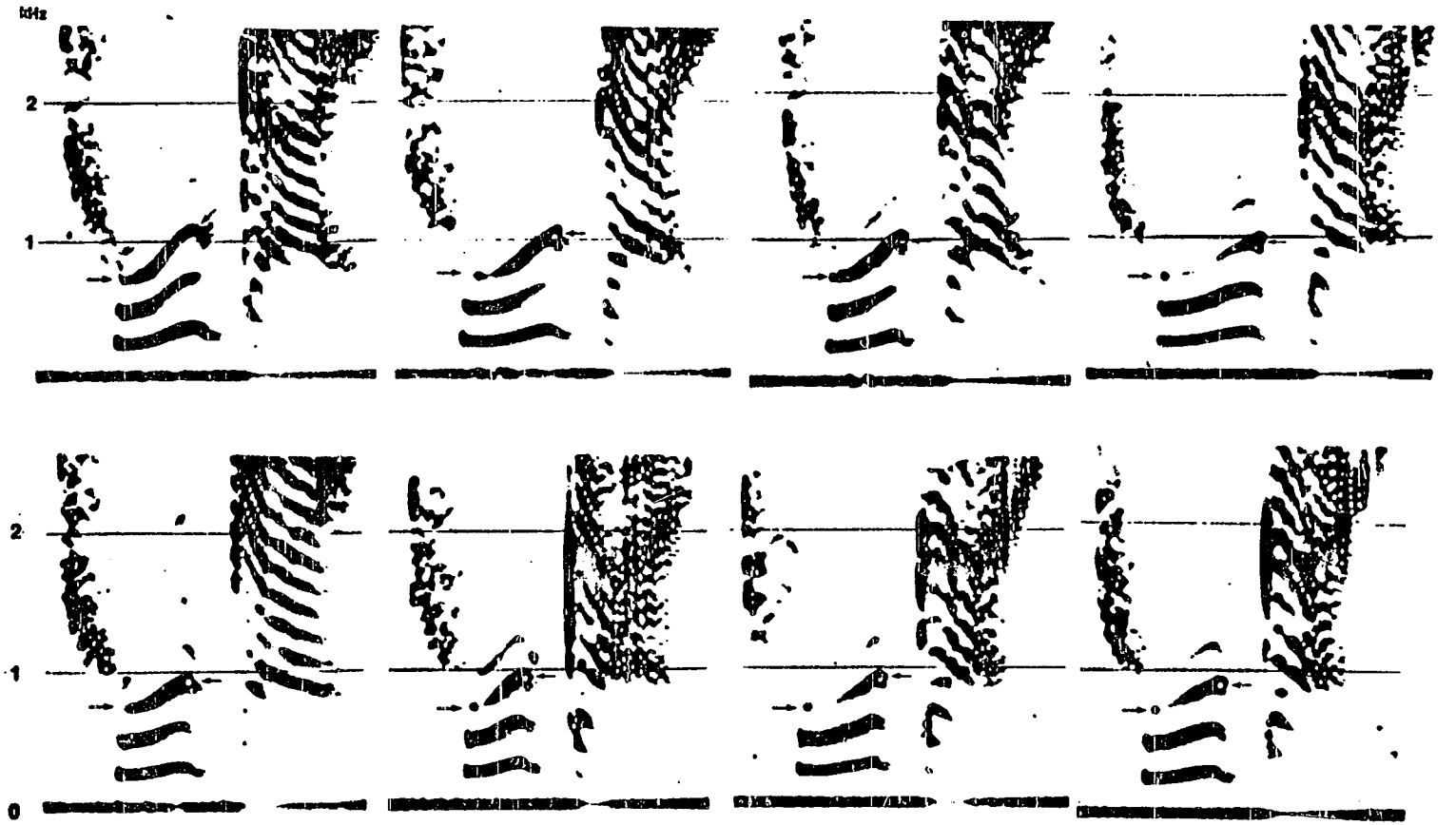


Figure 1a. Spectrograms of the bisyllabic words /tʰu kaɪ/ (35+214) 'to retouch' (the upper four) and /tʰu kaɪ/ (214+214) 'land re-distribution' (the lower four) for Speaker Q.M.

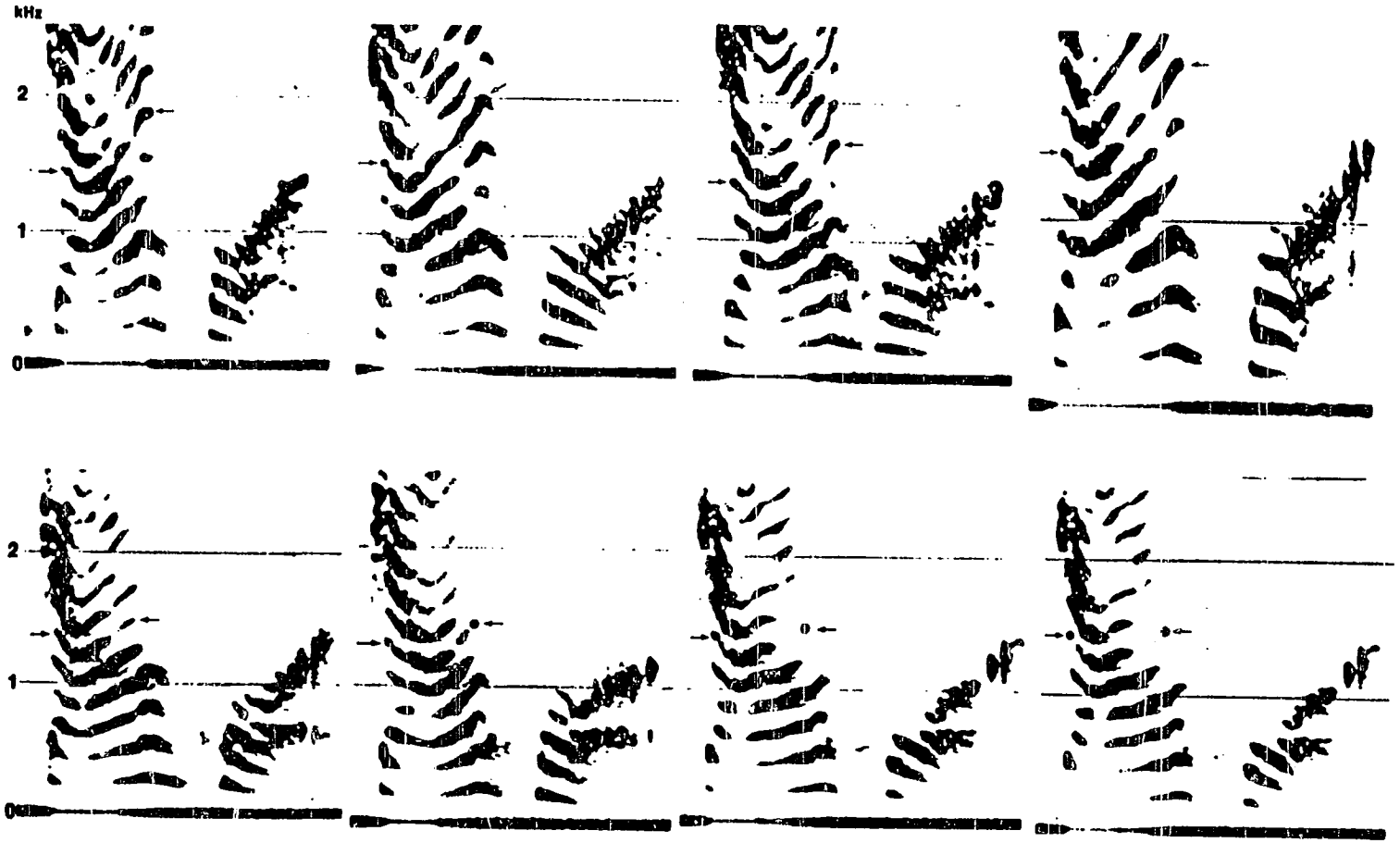


Figure 1b. Spectrograms of the bisyllabic words /tʂau xuo/ (35+214) 'on fire' (the upper four) and /tʂau xuo/ (214+214) 'to look for fire' (the lower four) for Speaker Q.M.

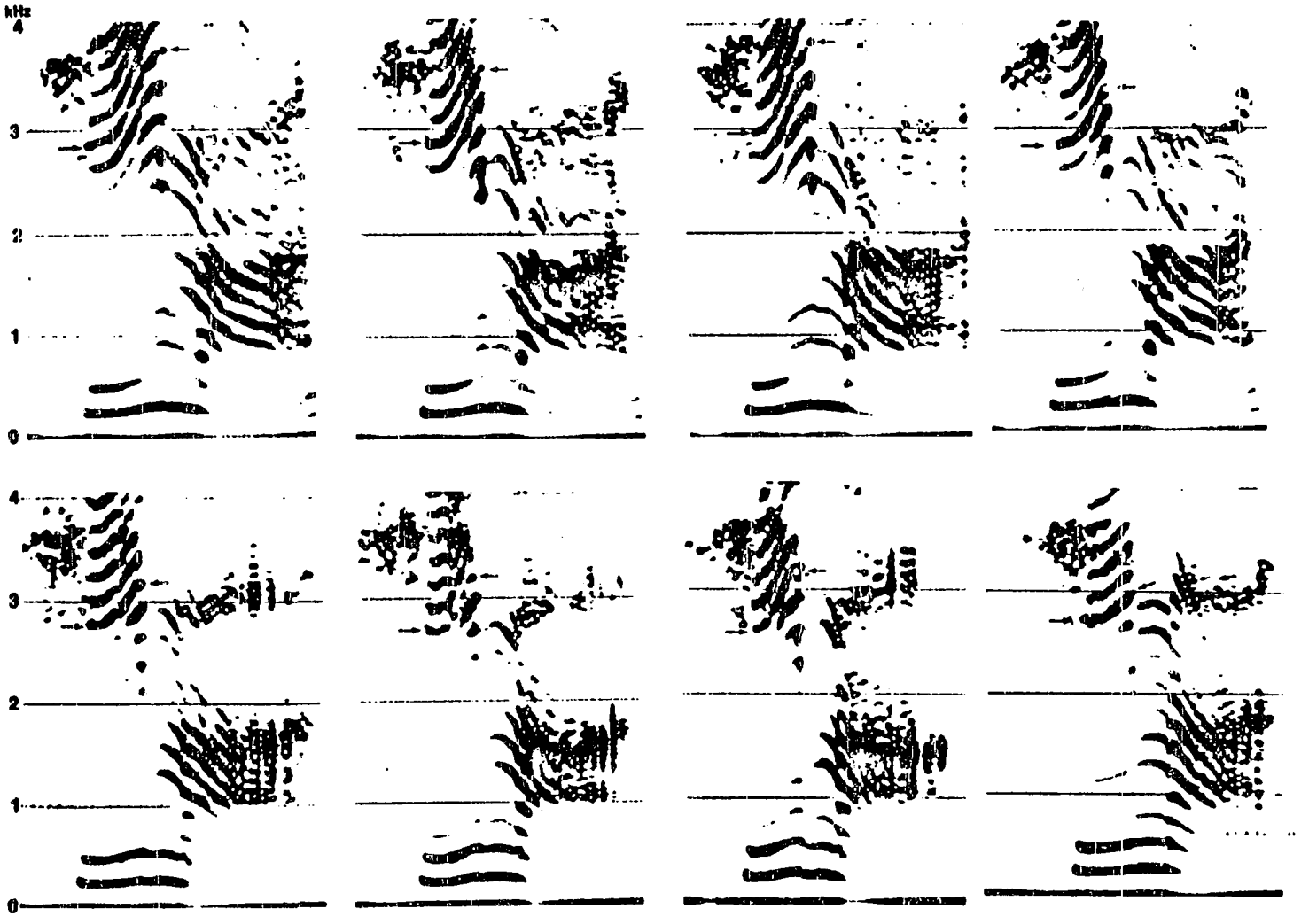


Figure 1c. Spectrograms of the bisyllabic words /tçʰi ma/ (35+214) 'to ride a horse' (the upper four) and /tçʰi ma/ (214+214) 'at least' (the lower four) for Speaker Q.M.

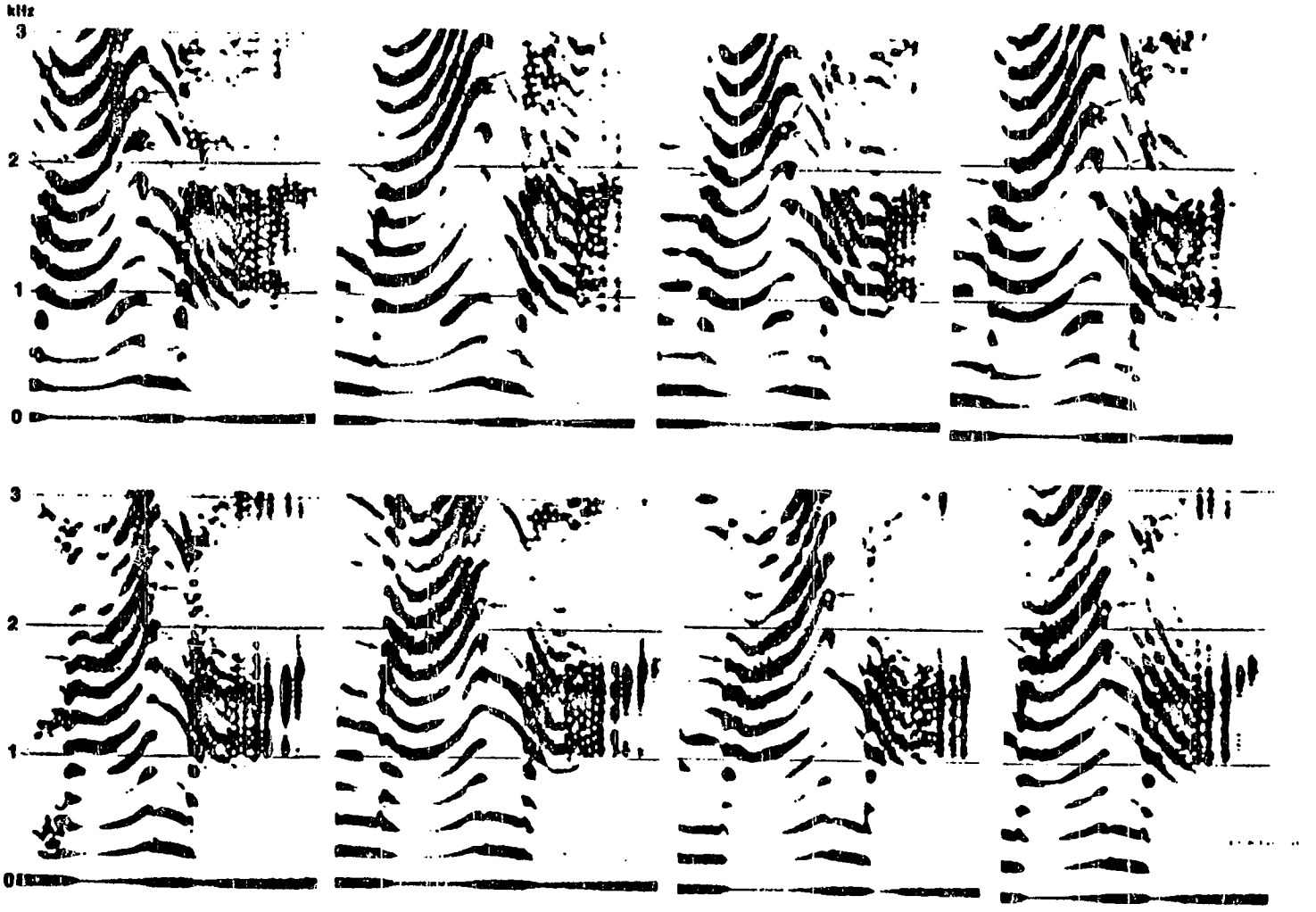


Figure 1d. Spectrograms of the bisyllabic words /mai ma/ (35+214) 'to bury a horse' (the upper four) and /mai ma/ (214+214) 'to buy a horse' (the lower four) for Speaker O.M.

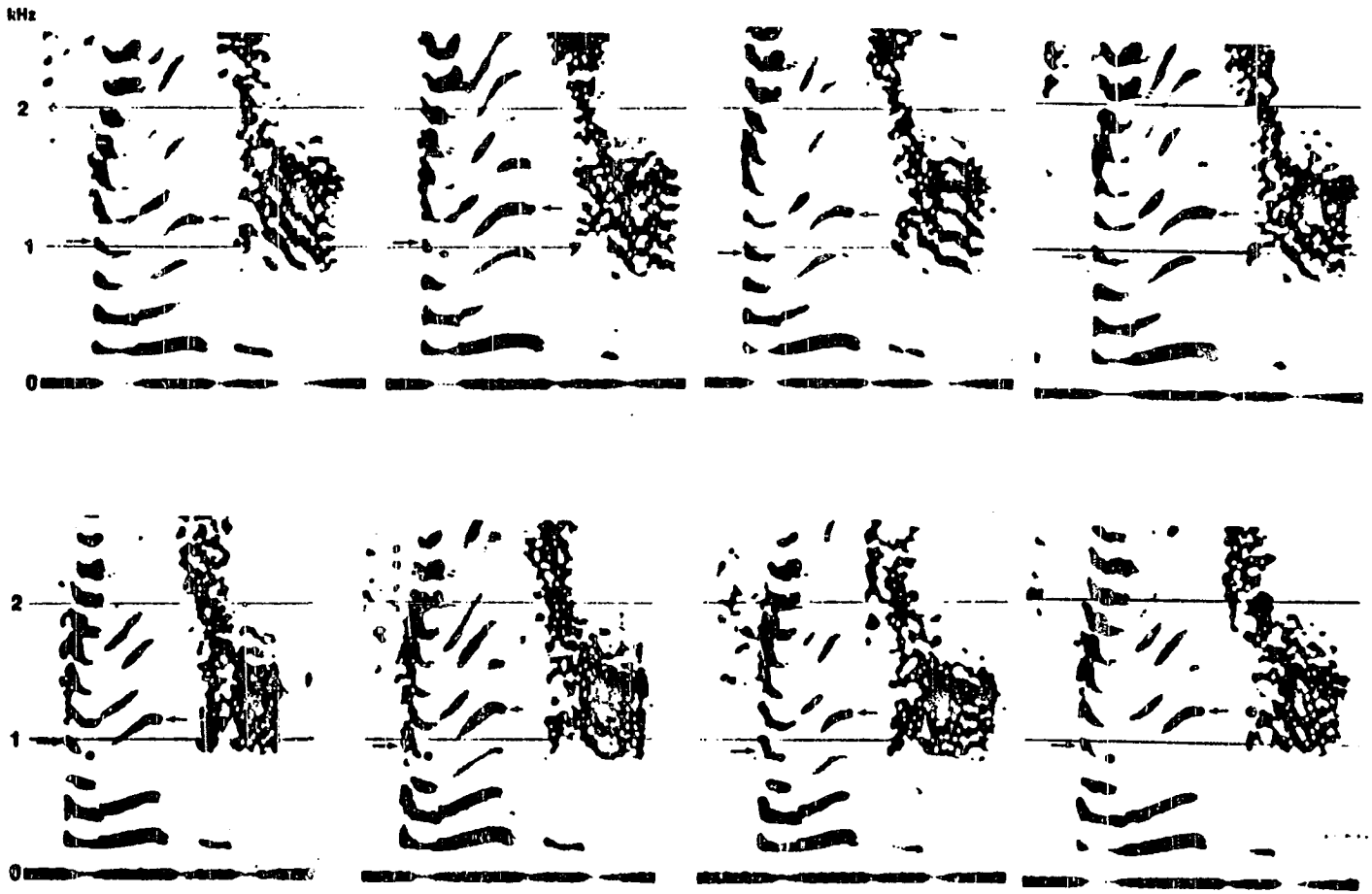


Figure 1e. Spectrograms of the bisyllabic words /fən tʂʰaŋ/ (35+214) 'graveyard' (the upper four) and /fən tʂʰan/ (214+214) 'a flour mill' (the lower four) for Speaker O.M.

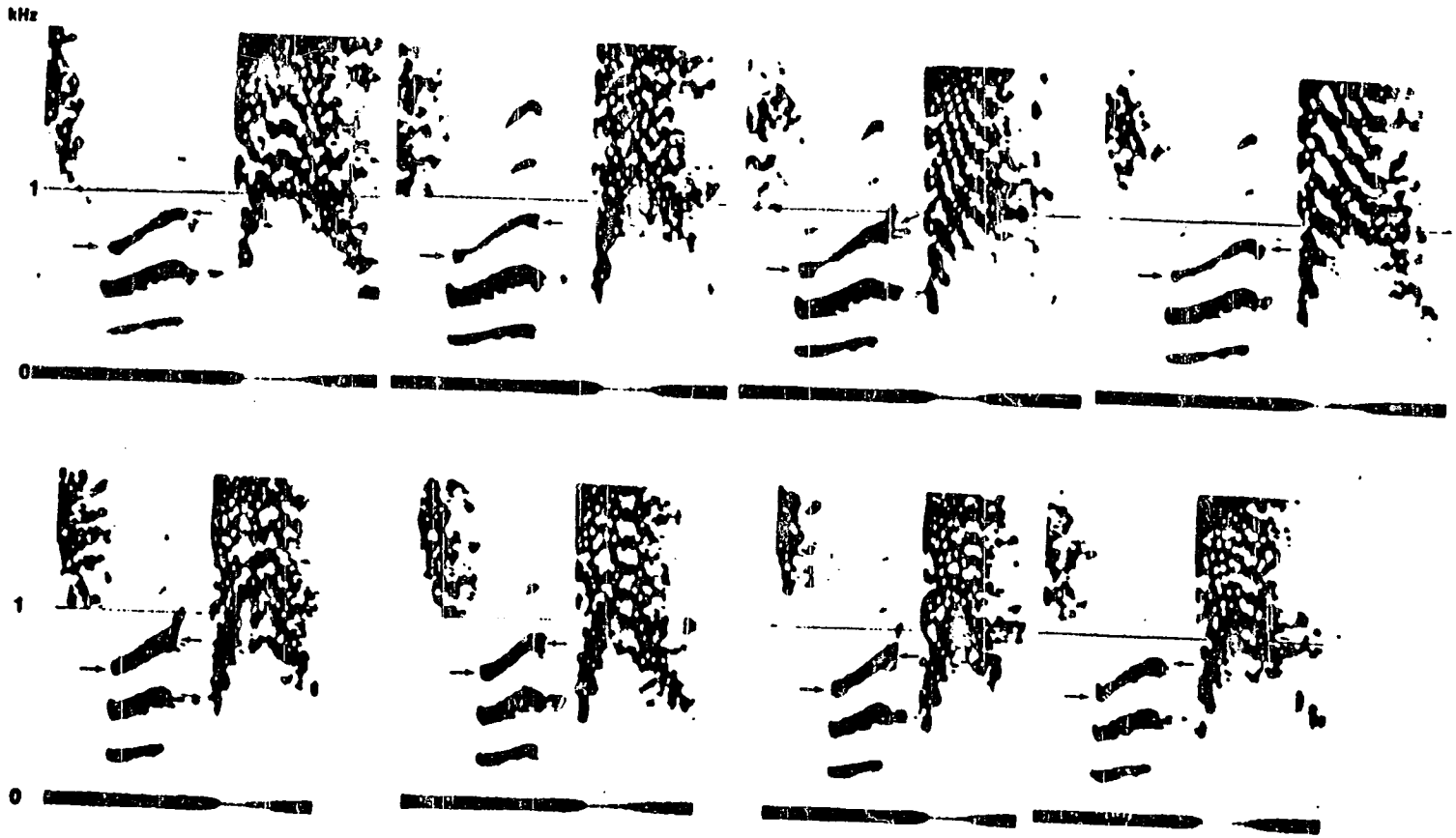


Figure 2a. Spectrograms of the bisyllabic words /tʰu kat/ (35+214) 'to retouch' (the upper four) and /tʰu kat/ (214+214) 'land re-distribution' (the lower four) for Speaker Y.H.J.

Figure 2b. Spectrograms of the bisyllabic words /tʂau xuo/ (35+214) 'on fire' (the upper four) and /tʂau xuo/ (214+214) 'to look for fire' (the lower four) for Speaker Y.H.J.

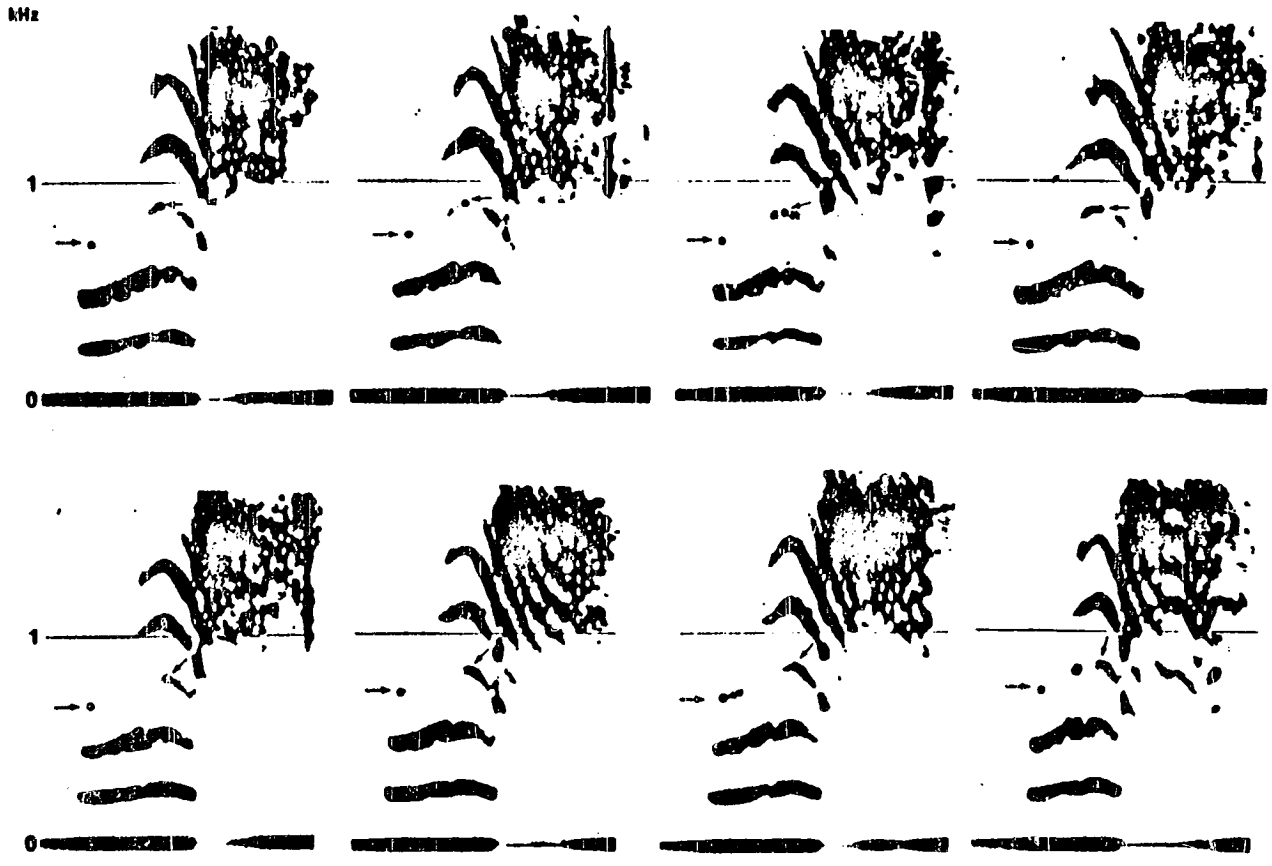


Figure 2c. Spectrograms of the bisyllabic words /tçʰi ma/ (35+214) 'to ride a horse' (the upper four) and /tçʰi ma/ (214+214) 'at least' (the lower four) for Speaker Y.H.J.

Figure 2d. Spectrograms of the bisyllabic words /mai ma/ (35+214) 'to bury a horse' (the upper four) and /mai ma/ (214+214) 'to buy a horse' (the lower four for Speaker Y.H.J.

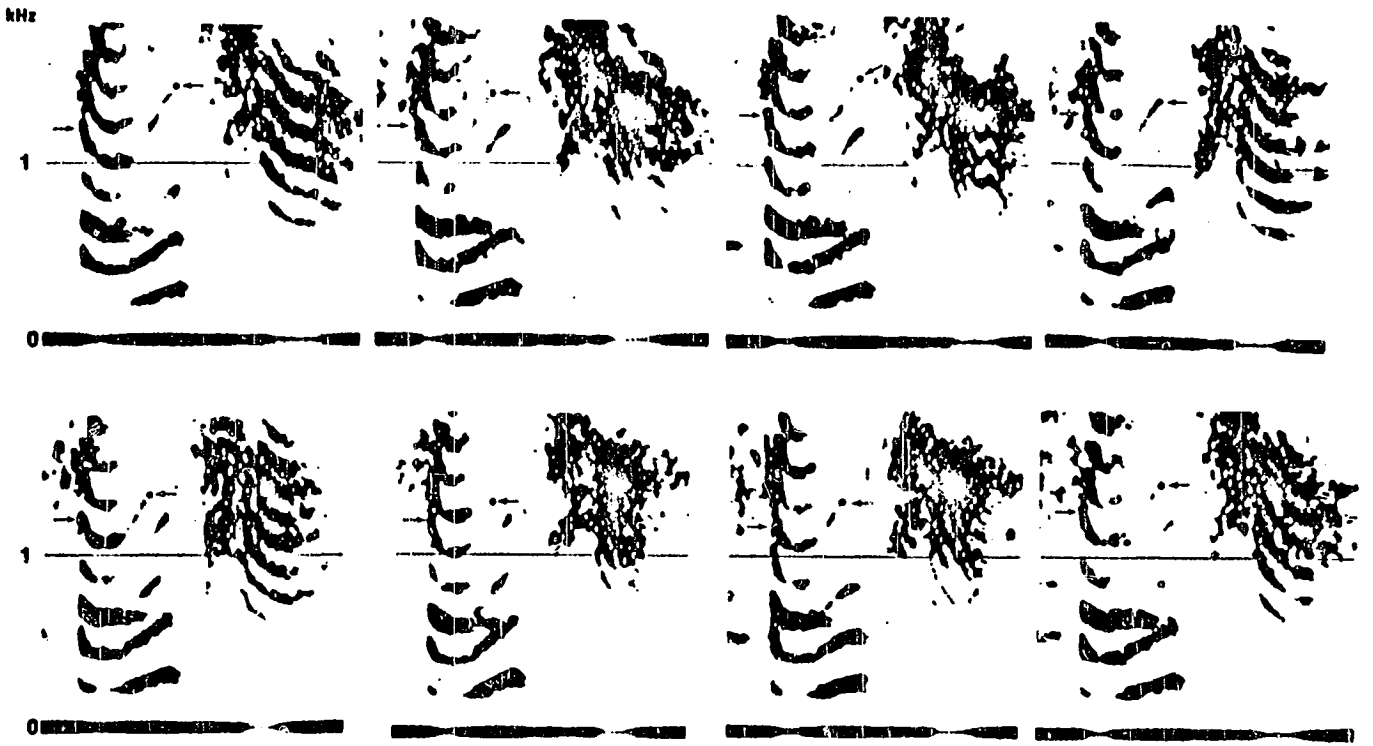


Figure 2a. Spectrograms of the bisyllabic words /fen tʂʰaŋ/ (35+214) 'graveyard' (the upper four) and /fen tʂʰaŋ/ (214+214) 'flour factory' (the lower four) for Speaker Y.H.J.

	I		II	
	[tʰu] in /tʰu+kaɪ/ (214+214) 'land redistribution'	[tʰu] in /tʰu+kaɪ/ (35+214) 'to re-touch'	[tʂau] in /tʂau+xuo/ (214+214) 'to look for fire'	[tʂau] in /tʂau+xuo/ (35+214) 'on fire'
	Beg. End	Beg. End	Beg. Dip End	Beg. Dip End
Speaker Q.M.	244.0/306.2	238.3/360.6	229.9/201.4/248.3	236.7/216.7/310.0
	245.7/313.3	238.3/340.1	218.9/190.2/241.8	253.3/226.7/363.3
	231.5/306.4	243.6/333.0	234.7/204.7/241.5	236.7/203.3/286.7
	241.7/320.1	241.7/326.7	223.3/203.3/245.3	240.0/208.3/306.4
(X̄)	240.9/311.5	240.5/340.1	226.7/199.9/244.4	241.7/213.8/316.6
Speaker Y.H.J.	235.3/289.2	232.3/297.9	235.3/180.4/270.6	247.1/207.8/290.2
	238.8/293.5	230.4/299.0	240.8/180.3/274.6	236.0/200.0/284.0
	228.9/283.6	222.2/297.9	239.2/187.7/269.4	244.0/200.0/302.0
	227.3/282.8	230.8/292.3	232.0/180.0/270.9	244.8/200.0/297.9
(X̄)	232.6/287.3	229.0/296.8	236.8/182.1/271.4	243.0/202.0/293.5

Table Ia. F0 values for the beginning, (dip) and end points of the pitch contour on the 1st syllable in the bisyllabic words: /tʰu kaɪ/ (214+214) 'land re-distribution', /tʰu kaɪ/ (35+214) 'to retouch', /tʂau xuo/ (214+214) 'to look for fire', and /tʂau xuo/ (35+214) 'on fire' for both Speakers Q.M. and Y.H.J.

	III		IV	
	[tçʰi] in /tçʰi+ma/ (214+214) 'at least'	[tçʰi] in /tçʰi+ma/ (35+214) 'to ride a horse'	[mai] in /mai+ma/ (214+214) 'to buy a horse'	[mai] in /mai+ma/ (35+214) 'to bury a horse'
	Beg. End	Beg. End	Beg. Dip End	Beg. Dip End
Speaker Q.M.	231.3/264.6	239.4/315.0	216.8/211.7/283.2	234.7/227.1/316.3
	222.9/266.7	237.5/298.3	229.6/204.1/275.2	227.1/221.9/331.6
	218.8/264.6	245.8/318.8	213.2/203.4/279.4	250.0/237.2/290.8
	230.8/248.0	237.2/279.9	218.7/213.9/262.0	228.4/221.1/302.9
	(X̄) 226.0/261.0	240.0/303.0	219.5/208.3/275.0	235.1/226.8/310.4
Speaker Y.H.J.	226.8/268.5	240.1/297.8	220.0/198.0/268.0	220.4/200.0/277.5
	243.3/277.3	252.3/301.8	212.2/200.0/269.4	220.0/206.0/288.0
	234.2/279.3	240.8/283.8	216.0/200.0/265.5	216.3/208.2/292.2
	242.0/280.1	232.9/292.3	208.0/176.0/260.0	220.0/200.0/280.0
	(X̄) 236.6/276.3	241.5/293.9	214.1/193.5/265.7	219.2/203.6/284.0

Table Ib. F_0 values for the beginning, (dip) and end points of the pitch contour on the 1st syllable in the bisyllabic words: /tçʰi ma/ (214+214) 'at least', /tçʰi ma/ (35+214) 'to ride a horse', /mai ma/ (214+214) 'to buy a horse', and /mai ma/ (35+214) 'to bury a horse' for both Speakers Q.M. and J.H.J.

		V	
		[fən] in /fən+tʂʰaŋ/ (214+214) 'flour factory'	[fən] in /fən+tʂʰaŋ/ (35+214) 'graveyard'
		Beg. Dip End	Beg. Dip End
Speaker O.M.		244.8/221.9/291.7	260.2/239.8/301.0
		250.0/219.4/306.1	255.1/239.8/321.4
		244.9/219.4/295.9	240.0/229.6/316.3
		245.0/220.0/305.0	240.3/230.8/317.3
	($\bar{X}$)	246.2/220.2/299.7	248.9/235.0/314.0
Speaker Y.H.J.		239.9/216.7/268.7	238.7/200.0/283.9
		235.5/201.3/258.1	241.9/212.9/277.4
		244.8/214.2/269.7	258.1/214.8/295.5
		232.2/209.7/260.0	251.6/214.8/279.4
	($\bar{X}$)	238.1/210.5/264.1	247.6/210.6/284.5

Table Ic. F_0 values for the beginning, dip and end points of the pitch contour on the 1st syllable in the bisyllabic words: /fən tʂʰaŋ/ (214+214) 'flour factory' and /fən tʂʰaŋ/ (35+214) 'graveyard' for both Speakers O.M. and Y.H.J.

	Speaker Y.H.J.					Speaker Q.M.				
					($\bar{X}$)					($\bar{X}$)
[tʰu kaɪ]/214+214/	0.60	0.58	0.55	0.58	0.57	0.58	0.55	0.54	0.49	0.54
[tʰu kaɪ] /35+214/	0.66	0.68	0.61	0.65	0.65	0.53	0.50	0.55	0.55	0.53
[tʂau xuo]/214+214/	0.58	0.53	0.62	0.60	0.58	0.61	0.53	0.62	0.53	0.57
[tʂau xuo] /25+214/	0.66	0.65	0.62	0.68	0.65	0.48	0.55	0.52	0.53	0.52
[tçʰi ma]/214+214]	0.52	0.52	0.54	0.68	0.57	0.53	0.49	0.52	0.53	0.52
[tcʰi ma] /35+214/	0.60	0.60	0.61	0.60	0.60	0.50	0.49	0.52	0.50	0.50
[mai ma]/214+214/	0.62	0.61	0.61	0.53	0.59	0.50	0.49	0.48	0.53	0.50
[mai ma] /35+214/	0.60	0.55	0.56	0.55	0.57	0.55	0.44	0.48	0.47	0.49
[fen tʂʰaŋ]/214+214/	0.64	0.65	0.65	0.66	0.65	0.53	0.58	0.56	0.53	0.55
[fen tʂʰaŋ] /35+214/	0.67	0.67	0.72	0.76	0.71	0.59	0.62	0.55	0.53	0.57

Table Id. The durations (in second) of all the bisyllabic word by Spkrs. Q.M. and Y.H.J.

	Trisyllabic word	Lexical tones	1st Syllable Beg.	1st Syllable End	1st Syllable Diff.	2nd Syllable Beg.	2nd Syllable End	2nd Syllable Diff.	Pitch contour on the 2nd syllable
# 1.	[fən şuei liŋ]	(˥ ˧˩˧) 'watershed'	322.5	329.4	(+ 5.9)	310.4	294.1	(-16.3)	HIGH LEVEL
# 2.	[san nian tçi]	(˥ ˧˧) '3rd grade'	328.4	338.2	(+ 9.8)	328.4	318.6	(- 9.8)	
# 3.	[tuŋ nan fəŋ]	(˥ ˧˥) 'Southeast wind'	330.9	316.2	(-14.7)	315.0	310.0	(- 5.0)	MID LEVEL
# 4.	[tsʰuŋ lou piŋ]	(˥˧˩˧) 'onion oil cake'	354.9	326.9	(-28.0)	320.7	312.8	(- 7.9)	
# 5.	[çian ʐən tçaŋ]	(˥˧˩˧) 'cactus'	319.0	312.8	(- 6.2)	317.3	298.1	322.1 (-19.2)(+24.0)	DIPPING
# 6.	[tuŋ xɤ ian]	(˥ ˧˩) 'East Riverside'	326.9	346.2	(+19.3)	284.0	308.0	(+24.0)	RISING
# 7.	[çi ian şən]	(˥ ˧˥) 'American ginseng'	331.9	310.9	(-21.0)	297.8	316.2	(+18.4)	
# 8.	[şuei nəŋ fei]	(˧ ˧˥) 'Who can fly?'	283.9	296.3	(+12.4)	319.4	328.7	(+ 9.3)	
# 9.	[mei ian faŋ]	(˧ ˧˥) '(a personal name)'	271.6	289.8	(+18.2)	296.3	310.0	(+13.7)	
# 10.	[xai mei uan]	(˧ ˧˧) 'not yet finished'	274.0	307.7	(+33.7)	328.0	333.0	(+ 5.0)	
# 11.	[xan şu piau]	(˧ ˧˩˧) 'thermometer'	288.5	314.1	(+25.6)	317.5	336.5	(+19.0)	

Table II. F_0 values for the beginning and end points of the pitch of the 2nd (and the 1st) syllable in the trisyllabic words with the lexical tone on the 1st syllable being either HIGH LEVEL or MID-RISING for Speaker Q.M.

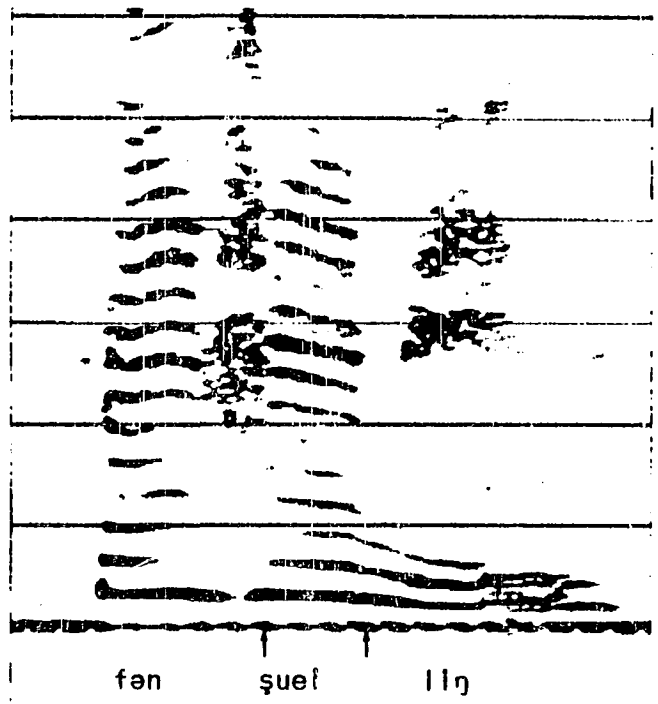


Figure 3.1 Spectrogram of /fən şuei lɪŋ/ (55+35+214) 'watershed' for Speaker Q.M.

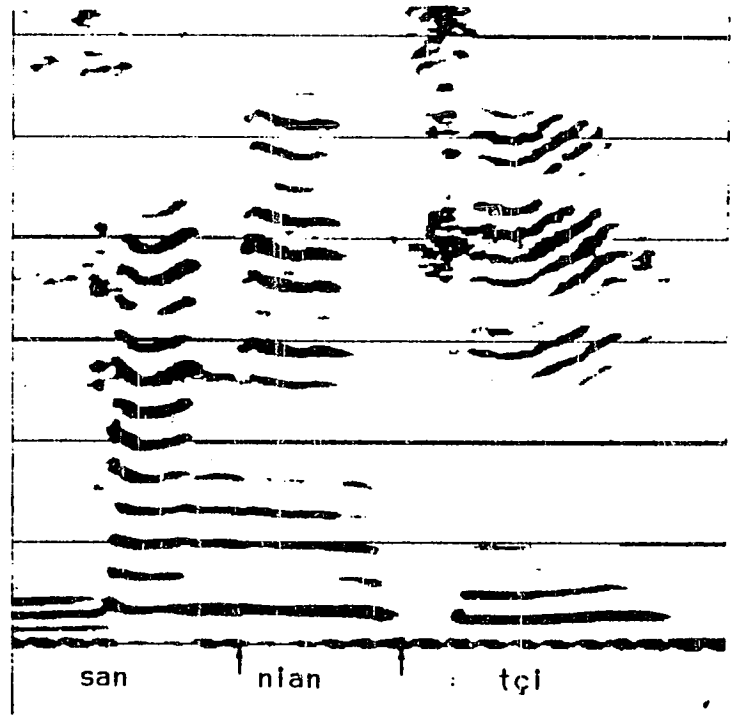


Figure 3.2 Spectrogram of /san nian tçi/ (55+35+35) '3rd grade' for Speaker Q.M.

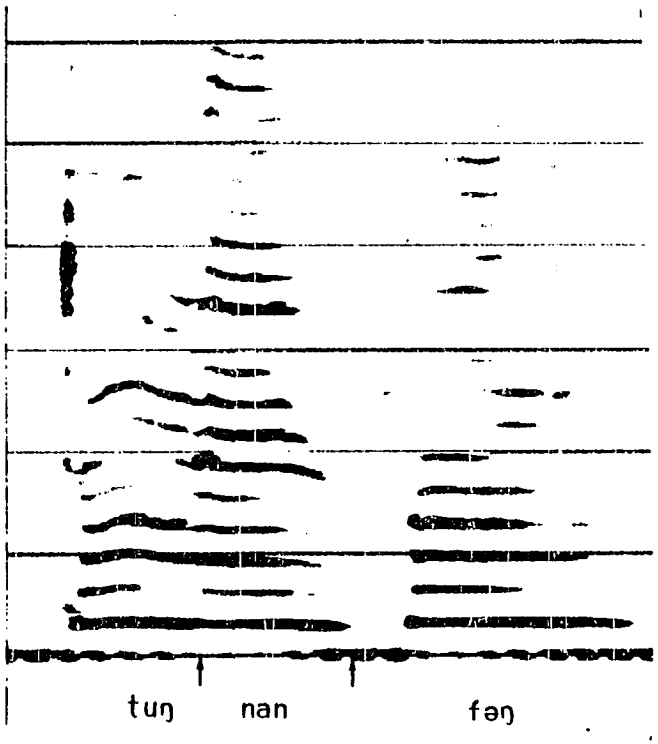


Figure 3.3 Spectrogram of /tuŋ nan fəŋ/ (55+35+55) 'Southeast wind' for Speaker Q.M.

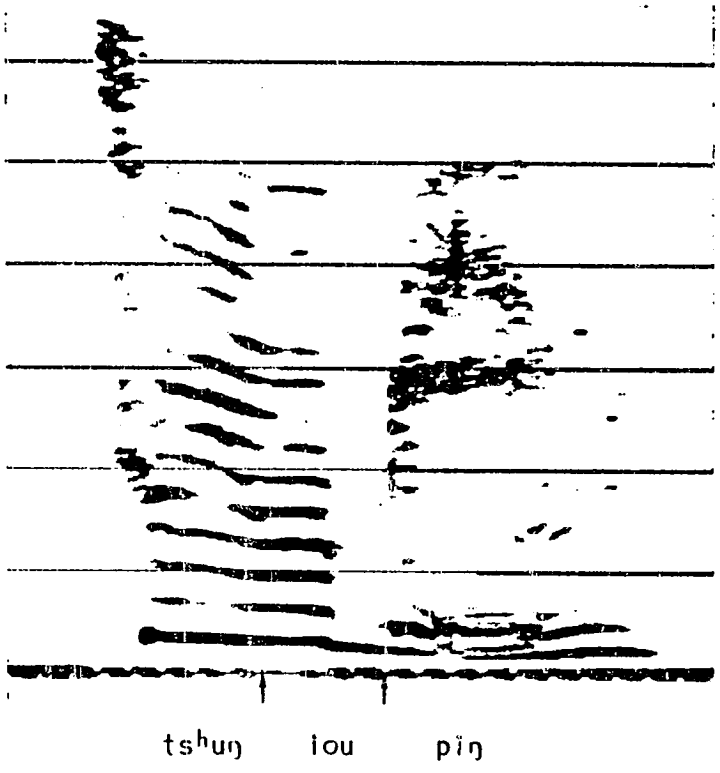


Figure 3.4 Spectrogram of /tsʰuŋ iou piŋ/ (55+35+214) 'onion oil cake' for for Speaker Q.M.

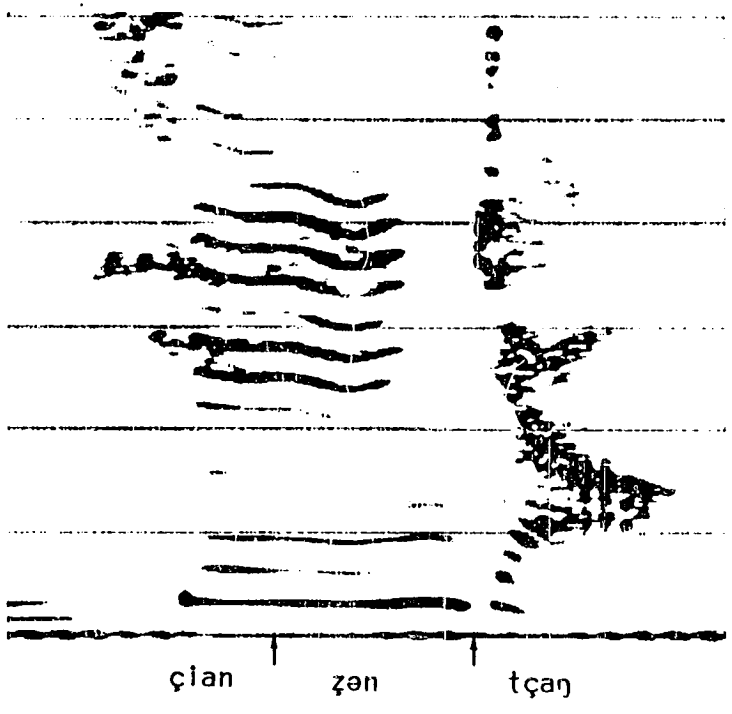


Figure 3.5 Spectrogram of /çian ʐen tçaŋ/ (55+35+214) 'cactus' for Speaker Q.M.

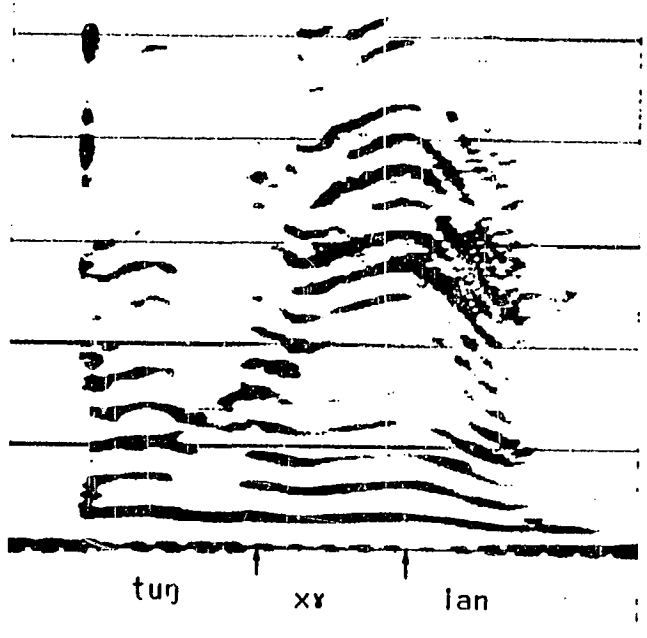


Figure 3.6 Spectrogram of /tuŋ xɤ lan/ (55+35+51) 'East Riverside' for Speaker Q.M.

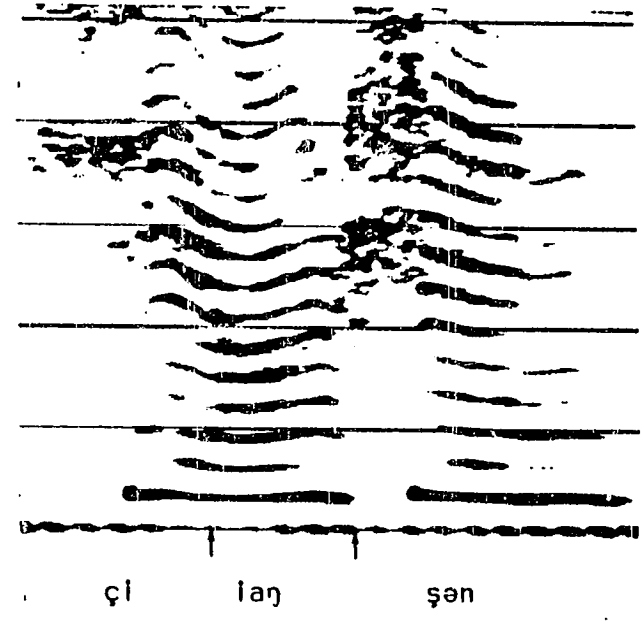


Figure 3.7 Spectrogram of /çi laŋ şən/ (55+35+55) 'American ginseng' for Speaker Q.M.

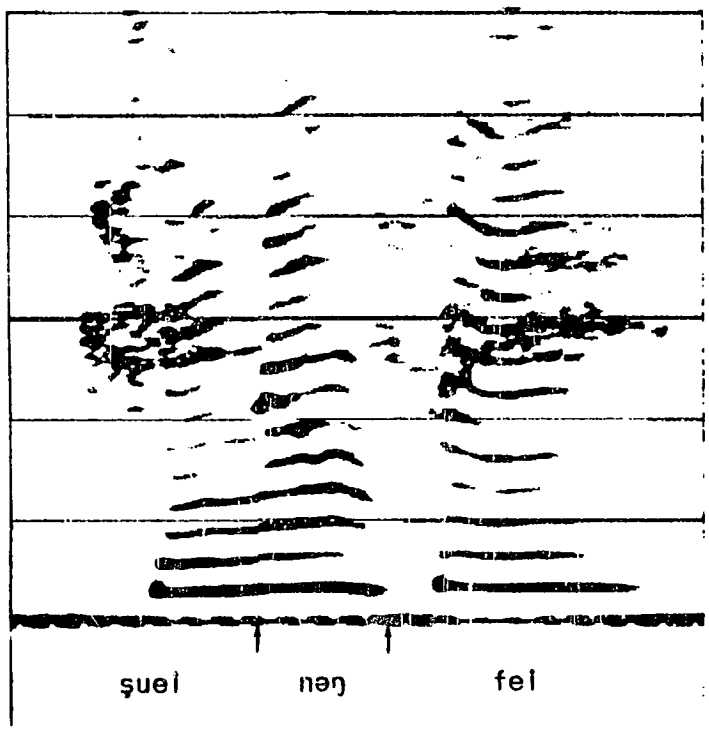


Figure 3.8 Spectrogram of /şuei nəŋ fei/ (35+35+55) 'Who can fly?' for for Speaker Q.M.

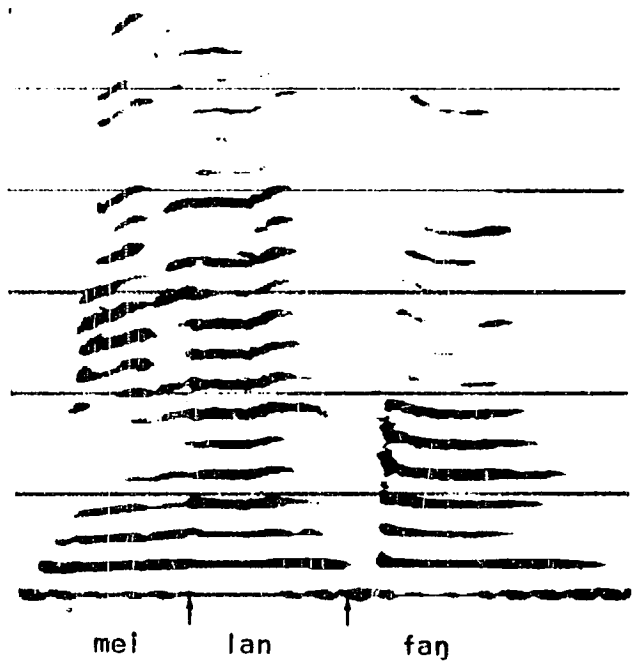


Figure 3.9 Spectrogram of /mei lan faŋ/ (35+35+55) '(a personal name)' for Speaker Q.M.

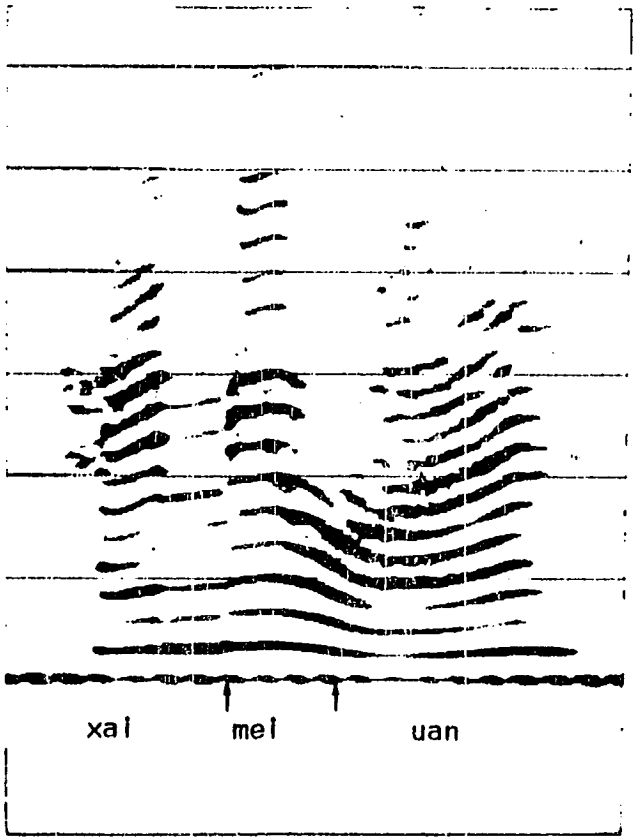


Figure 3.10 Spectrogram of /xai mei uan/ (35+35+35) 'not yet finished' for Speaker Q.M.

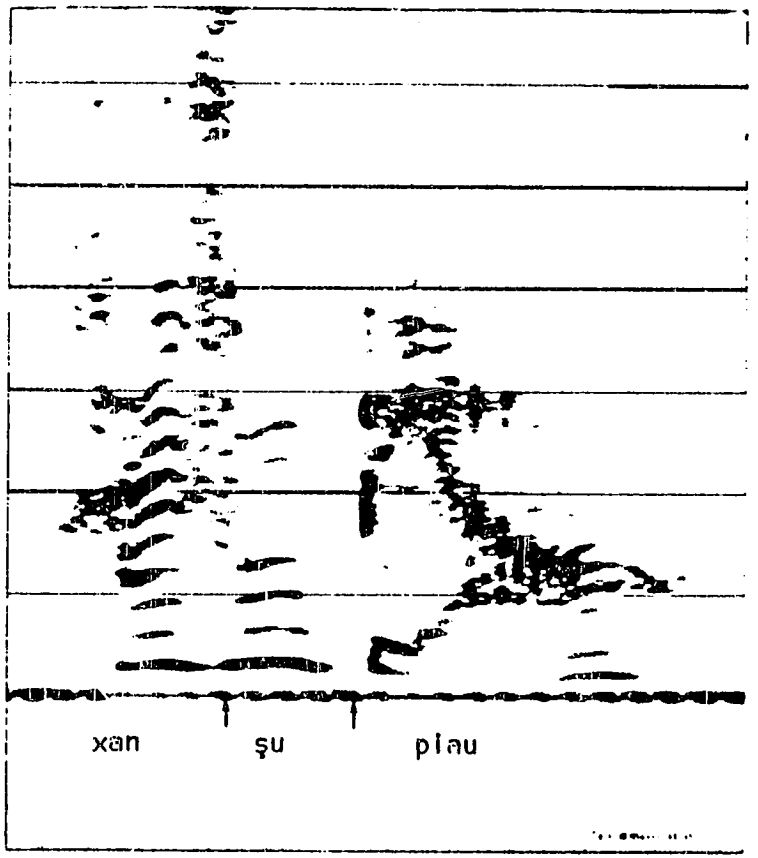


Figure 3.11 Spectrogram of /xan şu piau/ (35+35+214) 'thermometer' for Speaker Q.M.

	Trisyllabic word	Lexical tones	1st Syllable Beg.	1st Syllable End	1st Syllable Diff.	2nd Syllable Beg.	2nd Syllable End	2nd Syllable Diff.	Pitch contour on the 2nd syllable
# 1.	[xan şu piau] 'thermometer'	(˩˩˩)	201.5	224.1	(+22.6)	250.0	241.4	(- 8.6)	HIGH LEVEL
# 2.	[çi ian şən] 'American ginseng'	(˥˩˥)	241.0	206.0	(-35.0)	192.8	182.0	(-10.8)	MID LEVEL
# 3.	[tuŋ xɤ ian] 'East Riverside'	(˥˩˧)	259.3	258.0	(- 1.3)	226.0	210.0	(-16.0)	
# 4.	[çian ʐən tçaŋ] 'cactus'	(˥˩˩)	287.4	265.5	(-21.9)	235.7	211.1	(-24.6)	FALLING
# 5.	[san nian tçi] '3rd grade'	(˥˩˩)	229.9	235.6	(+ 5.7)	235.0	192.8	(-40.0)	
# 6.	[xai mei uan] 'not yet finished'	(˩˩˩)	202.4	234.2	(+31.8)	231.5	237.9	(+ 6.4)	RISING
# 7.	[fən şuei liŋ] 'watershed'	(˥˩˩)	250.0	255.0	(+ 5.0)	202.8	222.2	(+19.4)	
# 8.	[tuŋ nan fəŋ] 'Southeast wind'	(˥˩˥)	268.6	259.3	(- 9.3)	208.4/180.6/201.5	(-27.8)	(+20.9)	DIPPING
# 9.	[mei ian fəŋ] '(a personal name)'	(˩˩˥)	200.1	239.3	(+39.2)	230.2/202.4/222.3	(-27.8)	(+19.9)	
#10.	[tsʰuŋ iou piŋ] 'onion oil cake'	(˥˩˩)	255.6	241.4	(-14.2)	237.9/214.2/221.7	(-23.7)	(+ 7.5)	
#11.	[şuei nəŋ fei] 'Who can fly?'	(˩˩˥)	202.0	232.4	(+30.4)	205.2/197.6/203.7	(- 7.6)	(+ 6.0)	

Table III. F_0 values for the beginning and end points of the pitch of the 2nd (and the 1st) syllable in the trisyllabic words with the lexical tone on the 1st syllable being either HIGH LEVEL or MID-RISING for Speaker Y.H.J.

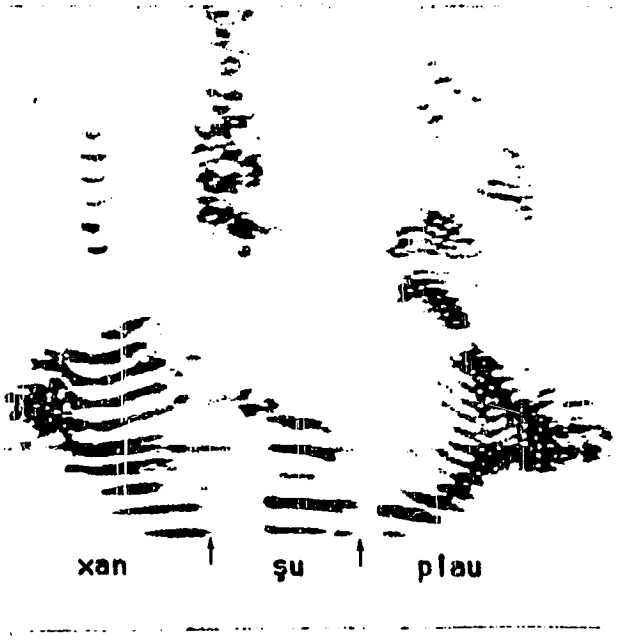


Figure 4.1 Spectrogram of /xan şu piau/ (35+35+214) 'thermometer' for Speaker Y.H.J.

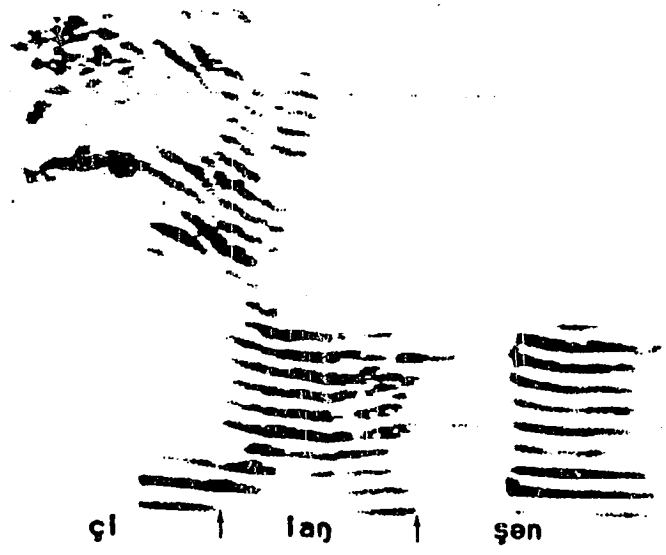


Figure 4.2 Spectrogram of /çɨ ɨaŋ şən/ (55+35+55) 'American ginseng' for Speaker Y.H.J.

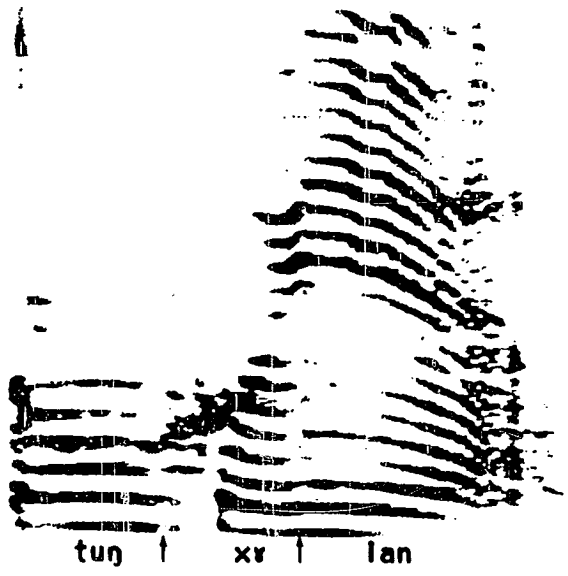


Figure 4.3 Spectrogram of /tuŋ xɤ ɨan/ (55+35+51) 'East Riverside' for speaker Y.H.J.

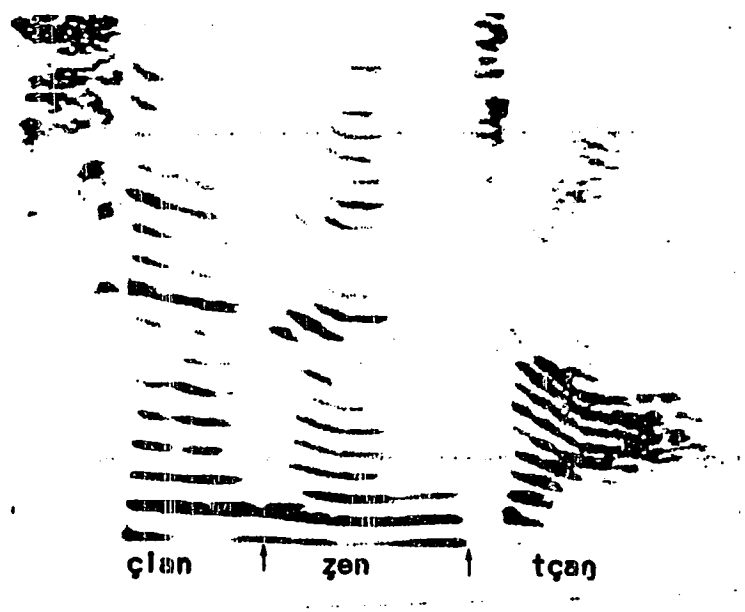


Figure 4.4 Spectrogram of /çlan ʐen tçaŋ/ (55+35+214) 'cactus' for Speaker Y.H.J.

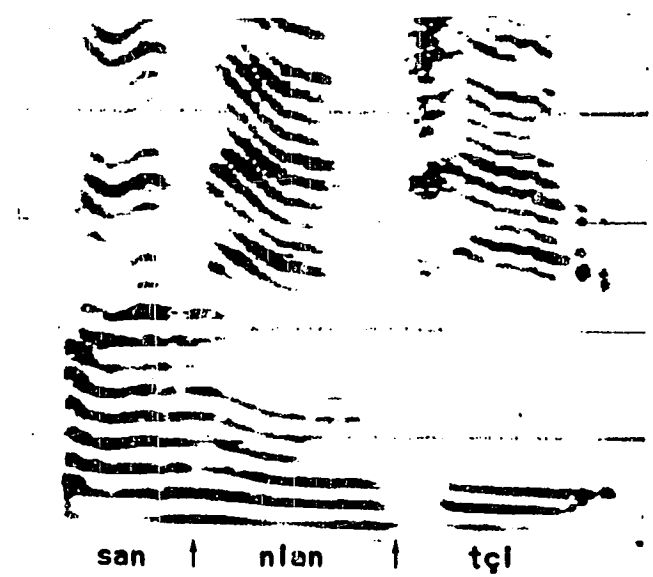


Figure 4.5 Spectrogram of /san nlan tçl/ (55+35+35) '3rd grade' for for Speaker Y.H.J.

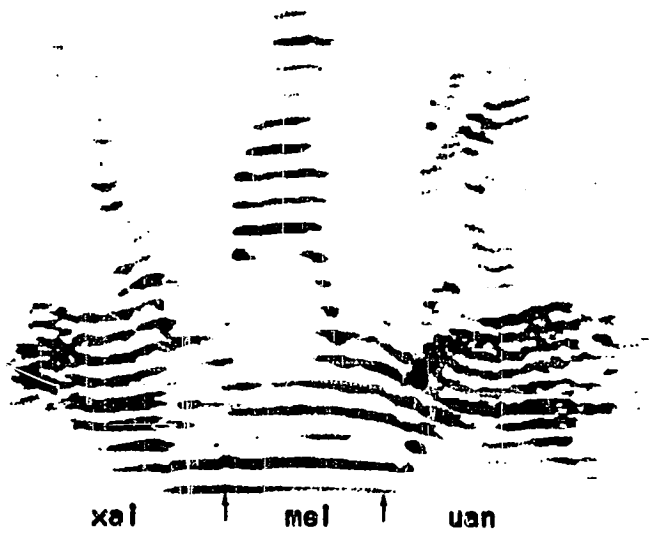


Figure 4.6 Spectrogram of /xaɪ meɪ uan/ (35+35+35) 'not yet finished' for Speaker Y.H.J.

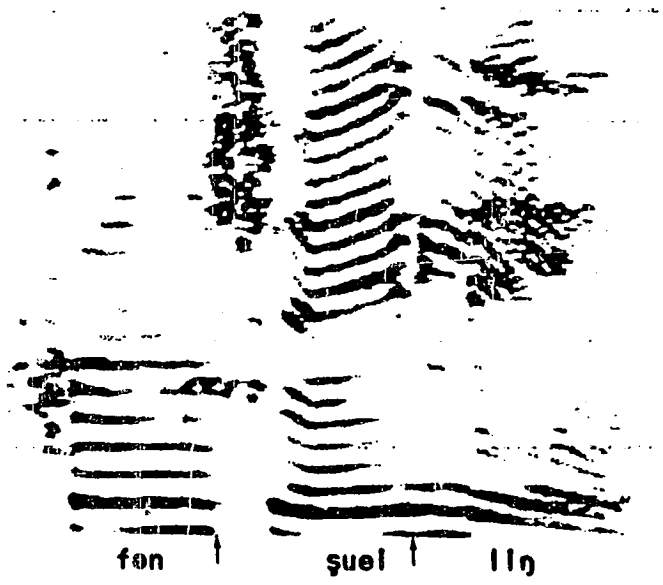


Figure 4.7 Spectrogram of /fən ʂueɪ lɪŋ/ (55+35+214) 'watershed' for Speaker Y.H.J.

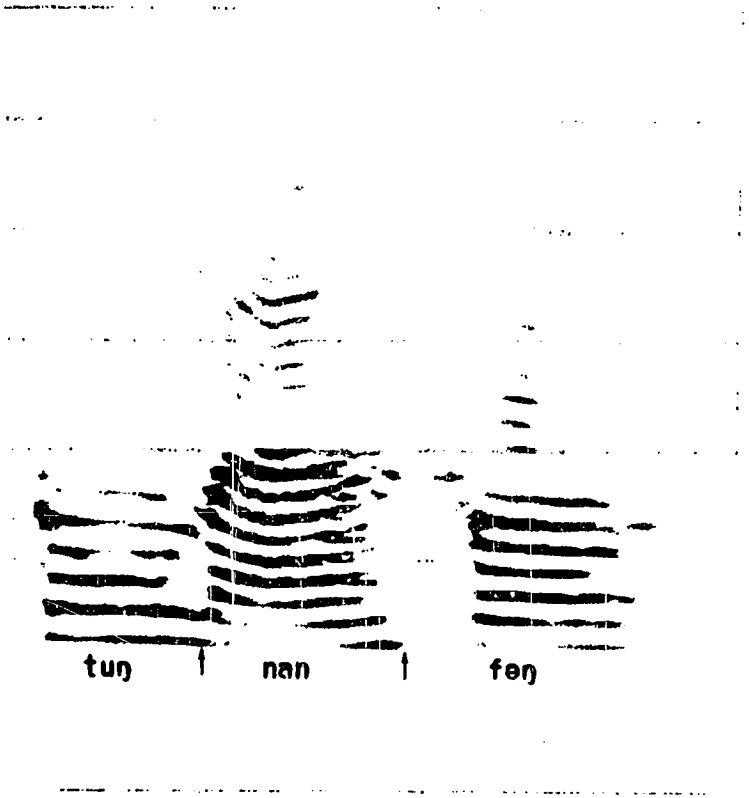


Figure 4.8 Spectrogram of /tuŋ nan feŋ/ (55+35+55) 'Southeast wind' for Speaker Y.H.J.

Figure 4.9 Spectrogram of /mei lan faŋ/ (35+35+55) '(a personal name)' for Y.H.J.

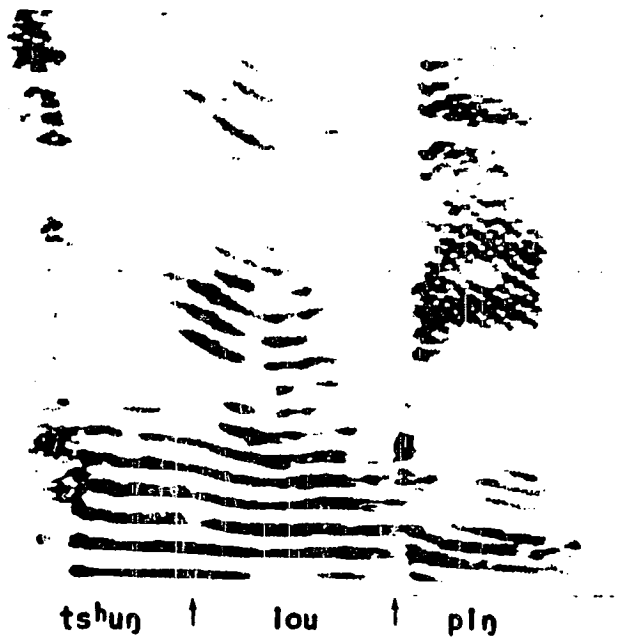


Figure 4.10 Spectrogram of /tsʰuŋ lou piŋ/ (55+35+214) 'onion oil cake' for Speaker Y.H.J.

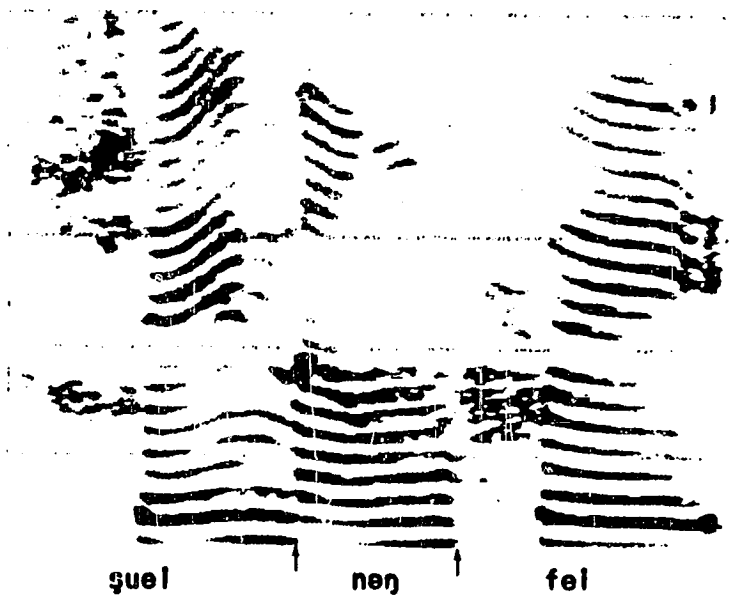


Figure 4.11 Spectrogram of /ʂuei neŋ fei/ (35+35+55) 'Who can fly?' for Speaker Y.H.J.

Trisyllabic words		Speaker Q.M.	Speaker Y.H.J.
[tuŋ nan fən]	'Southeast wind'	0.29	0.38
[çian ʐen tçaŋ]	'cactus'	0.38	0.43
[mei lan faŋ]	'(a personal name)'	0.31	0.38
[çi lan şən]	'American ginseng'	0.34	0.38
[san nian tçi]	'third grade'	0.36	0.38
[tsʰuŋ lou piŋ]	'onion oil cake'	0.29	0.38
[tuŋ xɤ lan]	'East Riverside'	0.32	0.29
[fən şuei liŋ]	'watershed'	0.26	0.41
[şuei nəŋ fei]	'Who can fly?'	0.29	0.43
[xai mei uan]	'not yet finished'	0.26	0.38
[xan şu piau]	'thermometer'	0.38	0.34
	($\bar{X}$)	0.32	0.38

Table IV. The total duration (in sec) for the first and second syllables in each of the trisyllabic words.

Spkr.	Speed	[lau] Beg.	[lau] End	[lau] Diff.	[lɪ] Beg.	[lɪ] End	[lɪ] Diff.	[mal] Beg.	[mal] End	[mal] Diff.	[çlau] Beg.	[çlau] End	[çlau] Diff.	Total duration
Q.M.	Fast	288.9	325.9	(+37.0)	308.0	209.9	(-98.0)	211.6	256.6	(+45.0)	236.1	259.3	(+23.2)	0.83 sec
	Fastest	259.3	327.2	(+67.9)	332.9	287.0	(-45.9)	259.3	281.5	(+22.2)	281.5	277.7	(- 3.8)	0.55 sec
Y.H.J.	Fast	203.7	270.0	(+66.3)	250.0	242.1	(- 7.9)	229.6	170.4	(-59.2)	192.6	237.1	(+44.5)	0.79 sec
	Fastest	204.6	222.2	(+17.6)	236.1	209.9	(-26.2)	200.0	181.9	(-18.1)	187.8	197.8	(+10.0)	0.53 sec

Table V. The F_0 values for the beginning and end points of the F_0 contour on each one of the first four syllables in the sentence [lau lɪ mal çlau pɪ] 'Old Li buys a small pen.' produced with two different speed by Speakers Q.M. and Y.H.J.

Spkr.	Speed	[lau]	[li]	[mai]	[çlau]
Q.M.	Fast	RISING	FALLING	RISING	RISING
	Fastest	RISING	FALLING	RISING	LEVEL
Y.H.J.	Fast	RISING	LEVEL	FALLING	RISING
	Fastest	RISING	FALLING	FALLING	RISING

Table VI. The shape of the F_0 contour on each one of the first four syllables in the sentence [lau li mai çlau pi] 'Old Li buys a small pen.' produced with two different speeds by Spkrs. Q.M. & Y.H.J.

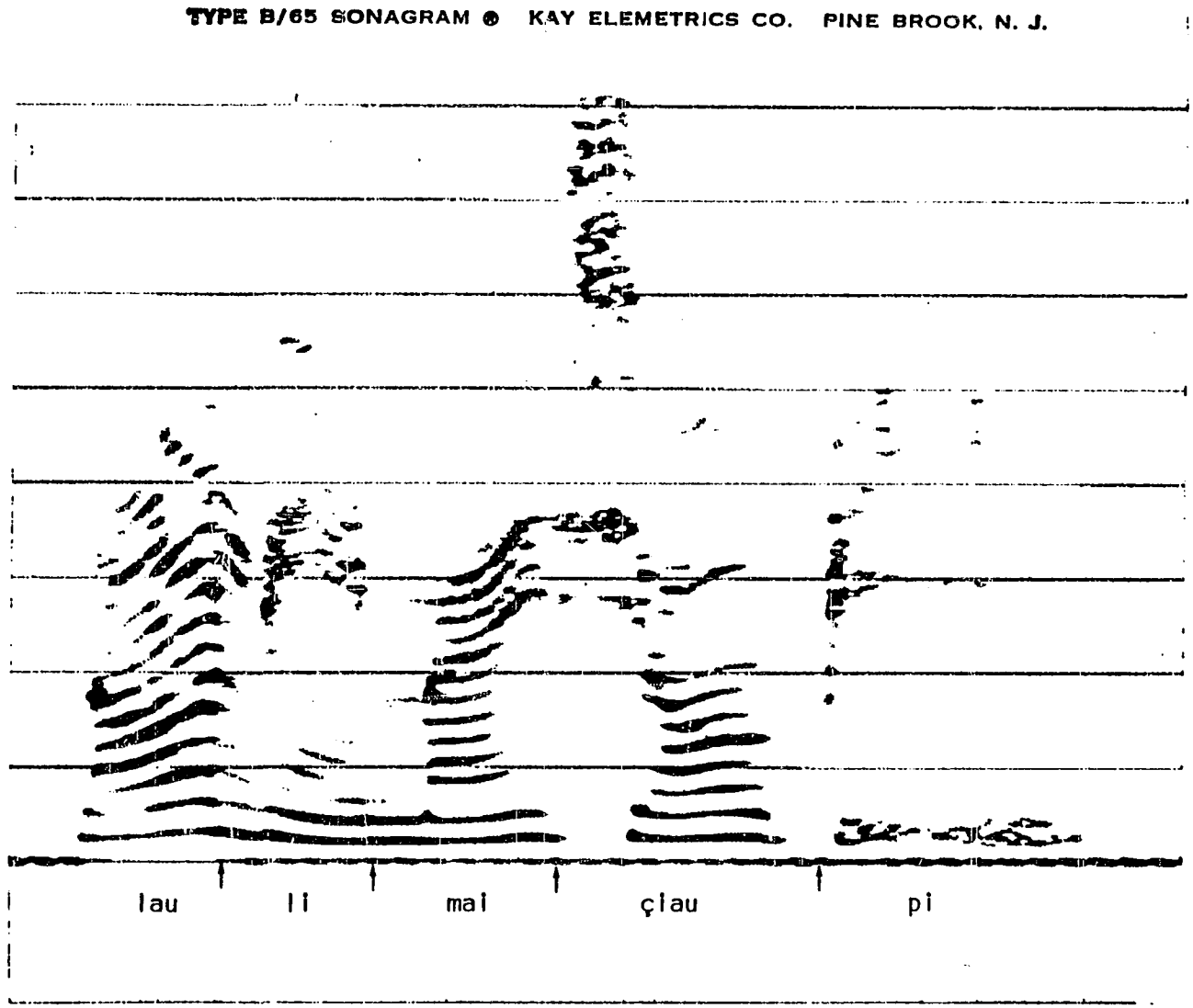


Figure 5a. Spectrogram of /lau li mai çiau pi/ (214+214+214+214+214) 'Old Li buys a small pen' in fast speech by Speaker Q.M.

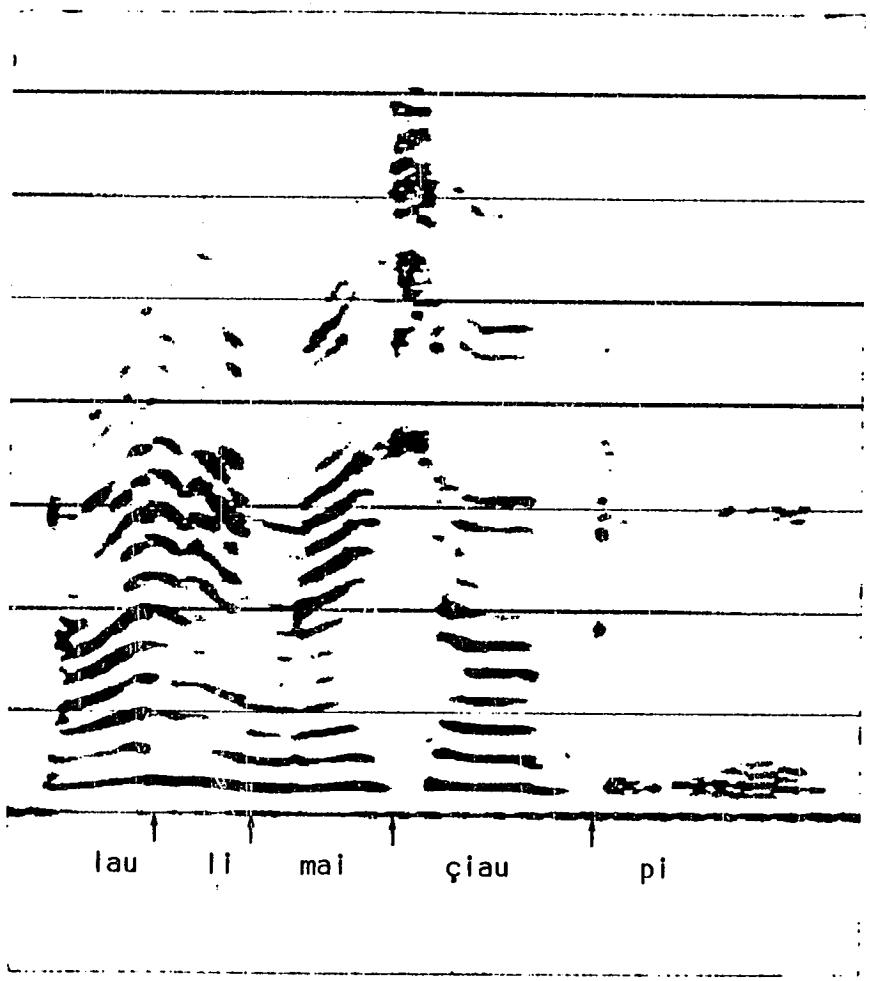


Figure 5b. Spectrogram of /lau li mai çiau pi/ (214+214+214+214+214) 'Old Li buys a small pen' in the fastest speech by Speaker Q.M.

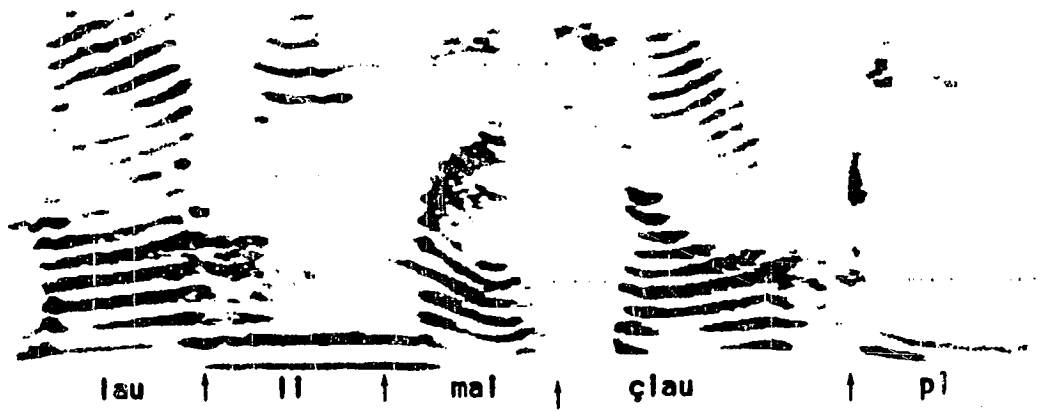


Figure 6a. Spectrogram of /lau li mai çiau pi/ (214+214+214+214+214) 'Old Li buys a small pen' in fast speech by Speaker Y.H.J.

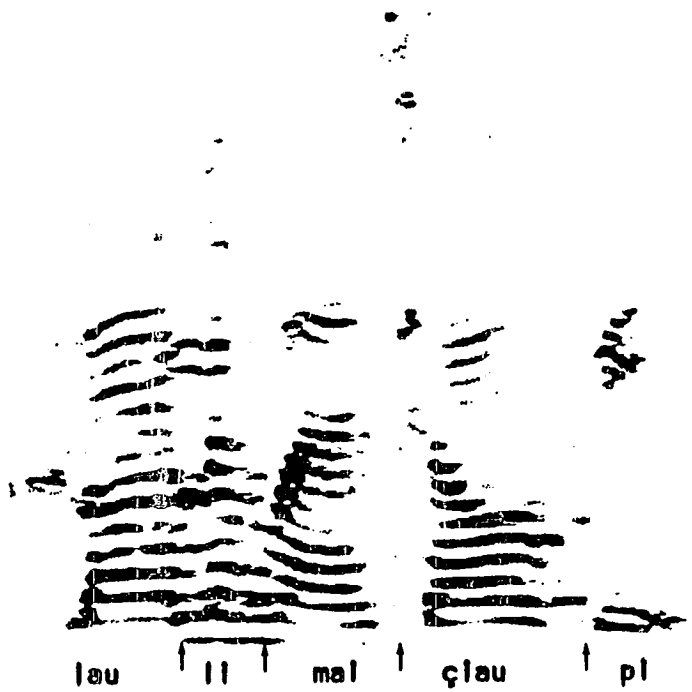


Figure 6b. Spectrogram of /lau li mai çiau pi/ (214+214+214+214+214) 'Old Li buys a small pen' in the fastest speech by Speaker Y.H.J.